

**Proceeding of the**

**20<sup>th</sup> European Conference on Cyber Warfare  
and Security**

**ECCWS 2021**

**A Virtual Conference**

**Hosted By**

**University of Chester**

**UK**

**24th-25th June 2021**

Copyright the authors, 2021. All Rights Reserved.

No reproduction, copy or transmission may be made without written permission from the individual authors.

### **Review Process**

Papers submitted to this conference have been double-blind peer reviewed before final acceptance to the conference. Initially, abstracts were reviewed for relevance and accessibility and successful authors were invited to submit full papers. Many thanks to the reviewers who helped ensure the quality of all the submissions.

### **Ethics and Publication Malpractice Policy**

ACIL adheres to a strict ethics and publication malpractice policy for all publications – details of which can be found here:

<http://www.academic-conferences.org/policies/ethics-policy-for-publishing-in-the-conference-proceedings-of-academic-conferences-and-publishing-international-limited/>

### **Self-Archiving and Paper Repositories**

We actively encourage authors of papers in ACIL conference proceedings and journals to upload their published papers to university repositories and research bodies such as ResearchGate and Academic.edu. Full reference to the original publication should be provided.

### **Conference Proceedings**

The Conference Proceedings is a book published with an ISBN and ISSN. The proceedings have been submitted to a number of accreditation, citation and indexing bodies including Thomson ISI Web of Science and Elsevier Scopus.

Author affiliation details in these proceedings have been reproduced as supplied by the authors themselves.

The Electronic version of the Conference Proceedings is available to download from DROPBOX <https://tinyurl.com/ECCWS21> Select Download and then Direct Download to access the Pdf file. Free download is available for conference participants for a period of 2 weeks after the conference.

The Conference Proceedings for this year and previous years can be purchased from <http://academic-bookshop.com>

E-Book ISBN: 978-1-912764-43-3

E-Book ISSN: 2048-8610

Book version ISBN: 978-1-912764-99-0

Book Version ISSN: 2048-8602

Published by Academic Conferences International Limited

Reading, UK

+44 (0) 118 324 6938

[www.academic-conferences.org](http://www.academic-conferences.org)

[info@academic-conferences.org](mailto:info@academic-conferences.org)



## Contents

Paper Title	Author(s)	Page No
Preface		v
Committee		vi
Biographies		vii
Keynote Outlines		
Research papers		
The PUF Commitment: Evaluating the Stability of SRAM-Cells	Pascal Ahr, Christoph Lipps and Hans Dieter Schotten	1
Asylum Seekers From Russia to Finland: A Hybrid Operation by Chance?	Kari Alenius	11
Antarctica and Cyber-Security: Useful Analogy or Exposing Limitations?	Shadi Alshdaifat, Brett van Niekerk and Trishana Ramluckan	18
Evasion of Port Scan Detection in Zeek and Snort and its Mitigation	Graham Barbour, André McDonald and Nenekazi Mkuzangwe	25
The Manifestation of Chinese Strategies Into Offensive Cyberspace Operations Targeting Sweden	Johnny Bengtsson and Gazmend Huskaj	35
The Evolution of Cyber Fraud in the Past Decade	George-Daniel Bobric	44
AI-Powered Defend Forward Strategy	Jim Chen	52
Global Military Machine Learning Technology Development Tracking and Evaluation	Long Chen and Jianguo Chen	61
Global Social Network Warfare on Public Opinion	Long Chen and Jianguo Chen	71
Serious Games for Cyber Security: Elicitation and Analysis of End-User Preferences and Organisational Needs	Sabarathinam Chockalingam, Coralie Esnoul, John Eidar Simensen and Fabien Sechi	80
Effectiveness of Covert Communication Channel Mitigation Across the OSI Model	Tristan Creek, Mark Reith and Barry Mullins	90
Deepfake Video Detection	Shankar Bhawani Dayal and Brett van Niekerk	100
A Shoestring Digital Forensic Cyber Range for a Developing Country	Jaco du Toit and Sebastian von Solms	110
A Strategy for Implementing an Incident Response Plan	Alexandre Fernandes, Adail Oliveira, Leonel Santos and Carlos Rabadão	120
Are Encrypted Protocols Really a Guarantee of Privacy?	Jan Fesl, Michal Konopa, Jiří Jelínek, Yelena Trofimova, Jan Janeček, Marie Feslová, Viktor Černý and Ivo Bukovsky	130
Targeting in All-Domain Operations: Choosing Between Cyber and Kinetic Action	Tim Grant and Harry Kantola	139
Computer Aided Diagnostics of Digital Evidence Tampering (CADET)	Babak Habibnia, Pavel Gladyshev and Marco Simioni	149
Weaknesses of IoT Devices in the Access Networks Used by People in Their Homes	Aarne Hummelholm	159

<b>Paper Title</b>	<b>Author(s)</b>	<b>Page No</b>
Cyber Security Analysis for Ships in Remote Pilotage Environment	Aarne Hummelholm, Jouni Pöyhönen, Tiina Kovanen and Martti Lehto	169
A Review of National Cyber Security Strategies (NCSS) Using the ENISA Evaluation Framework	Angela Jackson-Summers	178
Some Cybersecurity Governance Imperatives in Securing the Fourth Industrial Revolution	Victor Jaquire, Petrus Duvenage and Sebastian von Solms	187
Critical Infrastructure Protection: Employer Expectations for Cyber Security Education in Finland	Janne Jaurimaa, Karo Saharinen and Sampo Kotikoski	195
Digital Forensic Readiness Implementation in SDN: Issues and Challenges	Nickson Karie and Craig Valli	203
Cyber Wargaming on the Strategic/Political Level: Exploring Cyber Warfare in a Matrix Wargame	Thorsten Kodalle	212
Cyber-Threat Analysis in the Remote Pilotage System	Tiina Kovanen, Jouni Pöyhönen and Martti Lehto	221
Impact of AI Regulations on Cybersecurity Practitioners	Louise Leenen, Trishana Ramluckan and Brett van Niekerk	230
Is Hacking Back Ever Worth it?	Antoine Lemay and Sylvain Leblanc	239
EU Digital Sovereignty: A Regulatory Power Searching for its Strategic Autonomy in the Digital Domain	Andrew Liaropoulos	246
Mandatory Cybersecurity Training for all Space Force Guardians	Banks Lin, Mark Reith and Wayne Henry	253
The Challenges to Cybersecurity Education in Developing Countries: A Case Study of Kosovo	Arianit Maraj, Cynthia Sutherland and William Butler	260
Studying the Challenges and Factors Encouraging Girls in Cybersecurity: A Case Study	Arianit Maraj, Cynthia Sutherland and William Butler	269
IoT Security and Forensics: A Case Study	Erik David Martin, Iain Sutherland and Joakim Kargaard	278
Cybersecurity and local Government: Imperative, Challenges and Priorities	Mmalerato Masombuka, Marthie Grobler and Petrus Duvenage	285
KSA for Digital Forensic First Responder: A job Analysis Approach	Ruhama Mohammed Zain, Zahri Yunos, Nur Farhana Hazwani, Lee Hwee Hsiung and Mustaffa Ahmad	294
The Unrehearsed Boom in Education Automation, Amid COVID-19 Flouts, a Potential Academic Integrity Cyber Risks (AICR)!	Fredrick Ochieng Omogah	303
How Penetration Testers View Themselves: A Qualitative Study	Olav Opedal	314
Cyber Range: Preparing for Crisis or Something Just for Technical People?	Jani Päijänen, Karo Saharinen, Jarno Salonen, Tuomo Sipola, Jan Vykopal and Tero Kokkonen	322
Multiple-Extortion Ransomware: The Case for Active Cyber Threat Intelligence	Bryson Payne and Edward Mienie	331
Resilience Management Concept for Railways and Metro Cyber-Physical Systems	Jyri Rajamäki	337

<b>Paper Title</b>	<b>Author(s)</b>	<b>Page No</b>
Digital Evidence in Disciplinary Hearings: Perspectives From South Africa	Trishana Ramluckan, Brett van Niekerk and Harold Patrick	346
Security and Safety of Unmanned Air Vehicles: An Overview	Sérgio Ramos, Tiago Cruz and Paulo Simões	357
The Rising Power of Cyber Proxies	Janine Schmoltdt	369
Connected, Continual Conflict: Towards a Cybernetic Model of Warfare	Keith Scott	375
Emergency Response Model as a Part of the Smart Society	Jussi Simola, Martti Lehto and Jyri Rajamäki	382
Joint All-Domain Command and Control and Information Warfare: A Conceptual Model of Warfighting	Joshua Sipper	392
Defensive Cyber Deception: A Game Theoretic Approach	Abderrahmane Sokri	401
Using Semantic-Web Technologies for Situation Assessments of Ethical Hacking High-Value Targets	Sanjana Suresh, Rachel Fisher, Radha Patole, Andrew Zeyher and Thomas Heverin	407
Educating the Examiner: Digital Forensics in an IoT and Embedded Environment	Iain Sutherland, Huw Read and Konstantinos Xynos	416
Interdependence of Internal and External Security	Ilkka Tikanmäki and Harri Ruoslahti	425
The Host Nation Support for the International Cyber Operations	Maija Turunen	433
A GDPR Compliant SIEM Solution	Ana Vazão, Leonel Santos, Adail Oliveira and Carlos Rabadão	440
The Threat of Juice Jacking	Namosha Veerasamy	449
Status Detector for Fuzzing-Based Vulnerability Mining of IEC 61850 Protocol	Gábor Visky, Arturs Lavrenovs and Olaf Maennel	454
Mobile Phone Surveillance: An Overview of Privacy and Security Legal Risks	Murdoch Watney	462
<b>PHD Papers</b>		471
The Impact of GDPR Infringement Fines on the Market Value of Firms	Adrian Ford, Ameer Al-Nemrat, Seyed Ali Ghorashi and Julia Davidson	473
Side Channel Attacks and Mitigations 2015-2020: A Taxonomy of Published Work	Andrew Johnson	482
Sanctions and Cyberspace: The Case of the EU's Cyber Sanctions Regime	Eleni Kapsokoli	492
How the Civilian Sector in Sweden Perceive Threats From Offensive Cyberspace Operations	Joakim Kävrestad and Gazmend Huskaj	499
Aviation Sector Computer Security Incident Response Teams: Guidelines and Best Practice	Faith Lekota and Marijke Coetzee	507
Biocyberwarfare and Crime: A Juncture of Rethought	Xavier-Lewis Palmer, Ernestine Powell and Lucas Potter	517

<b>Paper Title</b>	<b>Author(s)</b>	<b>Page No</b>
Matters of Biocybersecurity With Consideration to Propaganda Outlets and Biological Agents	Xavier-Lewis Palmer, Ernestine Powell and Lucas Potter	525
Bio-Cyber Operations Inspired by the Human Immune System	Seyedali Pourmoafi and Stilianos Vidalis	534
Space Cyber Threats and Need for Enhanced Resilience of Space Assets	Jakub Pražák	542
e-Health as a Target in Cyberwar: Expecting the Worst	Samuel Wairimu	549
Talos: A Prototype Intrusion Detection and Prevention System for Profiling Ransomware Behaviour	Ashley Charles Wood, Thaddeus Eze and Lee Speakman	558
<b>Masters Research Papers</b>		569
The use of Neural Networks to Classify Malware Families	Theodore Drewes and Joel Coffman	571
Employing Machine Learning Paradigms for Detecting DNS Tunnelling	Jitesh Miglani and Christina Thorpe	580
Analysis of API Driven Application to Detect Smishing Attacks	Pranav Phadke and Christina Thorpe	588
Evolving Satellite Control Challenges: The Arrival of Mega-Constellations and Potential Complications for Operational Cybersecurity	Carl Poole, Mark Reith and Robert Bettinger	597
<b>Work In Progress Papers</b>		603
Inter-Process CFI for Peer/Reciprocal Monitoring in RISC-V-Based Binaries	Toyosi Oyinloye, Lee Speakman and Thaddeus Eze	605
Use of Blockchain Technologies Within the Creative Industry to Combat Fraud in the Production and (Re)Sale of Collectibles	Alexander Pfeiffer, Stephen Bezzina and Thomas Wernbacher <sup>1</sup>	611
Peer2Peer Communication via Testnet Systems of Blockchain Networks: A new Playground for Cyberterrorists?	Alexander Pfeiffer, Thomas Wernbacher and Stephen Bezzina	615
Ethics of Cybersecurity in Digital Healthcare and Well-Being of Elderly at Home	Jyri Rajamäki	619
ECHO Federated Cyber Range as a Tool for Validating SHAPES Services	Jyri Rajamäki and Harri Ruoslahti	623

## Preface

These proceedings represent the work of contributors to the 20th European Conference on Cyber Warfare and Security (ECCWS 2021), supported by University of Chester, UK on 24-25 June 2021. The Conference Co-chairs are Dr Thaddeus Eze University of Chester and Dr Lee Speakman, University of Salford and the Programme Chair is Dr Cyril Onwubiko from IEEE and Director, Cyber Security Intelligence at Research Series Limited.

ECCWS is a well-established event on the academic research calendar and now in its 20th year the key aim remains the opportunity for participants to share ideas and meet. The conference was due to be held at University of Chester, UK, but due to the global Covid-19 pandemic it was moved online to be held as a virtual event. The scope of papers will ensure an interesting conference. The subjects covered illustrate the wide range of topics that fall into this important and ever-growing area of research.

The opening keynote presentation is given by *Detective Inspector David Turner, and Detective Constable Michael Roberts* on the topic of *Policing the UK Cyber Space*. There will be a second keynote at 12:45 on Thursday presented by: Detective Constable Will Farrell, and Police Constable Phil Byrom on *CyberChoices – Helping young people choose the right and legal path*. The second day of the conference will open with an address by of the *Keith Terrill, and Louisa Murphy* speaking on *Current Cyber Crime Patterns and Trends - Covering the Traditional and Dark Webs*.

With an initial submission of 116 abstracts, after the double blind, peer review process there are 54 Academic research papers, 11 PhD research papers, 4 Masters research paper and 5 work-in-progress papers published in these Conference Proceedings. These papers represent research from Australia, Austria, Canada, China, Czech Republic, Estonia, Finland, Germany, Greece, India, Ireland, KENYA, Kosovo, Malaysia, Netherlands, Norway, Pakistan, Portugal, Romania, South Africa, Sweden, UK and USA.

We hope you enjoy the conference.

Dr Thaddeus Eze  
University of Chester  
UK  
June 2021

## **ECCWS Conference Committee**

*Dr. Mohd Faizal Abdollah, University Technical Malaysia Melaka, Malaysia; Dr William ("Joe") Adams, Univ of Michigan/Merit Network, USA; Dr. Tariq Ahamad, Prince Sattam Bin Abdulaziz University, Saudi Arabia; Prof Hamid Alasadi, Basra University, Iraq; Dr. Kari Alenius, University of Oulu, Finland; Prof. Antonios Andreatos, Hellenic Air Force Academy, Greece; Dr. Olga Angelopoulou, University of Warwick, UK; Faculty John Anohar, Full time academic, Higher Education dept; Dr. Leigh Armistead, Edith Cowan University, Australia; Johnnes Arreymbi, University of East London, UK; Dr. Hayretdin Bahsi, Tallinn University of Technology, Estonia; Prof Jorge Barbosa, Full time academic, Coimbra Polytechnic - ISEC; Dr. Darya Bazarkina, Sholokhov Moscow State Humanitarian University, Russia; Mr Robert Bird, Coventry University, UK; Prof. Matt Bishop, University of California at Davis, USA; Dr Radomir Bolgov, Saint Petersburg State University, Russia; Dr. Svet Braynov, University of Illinois at Springfield, USA; Prof. Larisa Breton, FullCircle Communications, LLC, USA; Dr Jim Chen, DoD National Defense University, USA; Dr Sabarathinam Chockalingam, , Institute for Energy Technology,; Bruce Christianson, University of Hertfordshire, UK; Dr. Maura Conway, Dublin City University, Ireland; Dr. Paul Crocker, Universidade de Beira Interior, Portugal; Prof. Tiago Cruz, University of Coimbra, Portugal; Dr. Christian Czosseck, CERT Bundeswehr (German Armed Forces CERT), Germany; Geoffrey Darnton, Bournemouth University, UK; Josef Demergis, University of Macedonia, Greece; Prof Patricio Domingues, Full time academic, Polytechnic Institute of Leiria; Paul Dowland, Edith Cowan University, Australia; Marios Efthymiopoulos, Political Science Department University of Cyprus, Cyprus; Dr. Colin Egan, University of Hertfordshire, Hatfield, UK; Dr Ruben Elamiryan, Public Administration Academy of the Republic of Armenia, Armenia; Prof. Dr. Alptekin Erkollar, ETCOP, Austria; Dr Thaddeus Eze, University of Chester, UK; John Fawcett, University of Cambridge, UK; Prof. Eric Filiol, ENSIBS, Vannes, France & CNAM, Paris, France; Dr. Chris Flaherty, University of New South Wales, Australia; Prof. Steve Furnell, University of Nottingham, UK; Mr. Tushar Gokhale, Hewlett Packard Enterprise, USA; Dr. Michael Grimaila, Air Force Institute of Technology, USA; Prof. Stefanos Gritzalis, University of the Aegean, Greece; Dr. Mils Hills, Northampton Business School, UK; Dr Ulrike Hugel, University of Innsbruck, Austria; Aki Huhtinen, National Defence College, Finland; Bill Hutchinson, Edith Cowan University, Australia; Dr. Abhaya Induruwa, Canterbury Christ Church University, UK; Hamid Jahankhani, University of East London, UK; Nor Badrul Anuar Jumaat, University of Malaya, Malaysia; Maria Karyda, University of the Aegean, Greece; Ass. Prof. Vasilis Katos , Democritus University of Thrace, Greece; Dr. Anthony Keane, Technological University Dublin, Ireland; Jyri Kivimaa, Cooperative Cyber Defence and Centre of Excellence, Tallinn, Estonia; Prof. Ahmet Koltuksuz, Yasar University, Dept. of Comp. Eng, Turkey; Dr Maximiliano Korstanje, Full time academic, University of Palermo, Buenos Aires, Argentina; Prashant Krishnamurthy, University of Pittsburgh, USA; Mr. Peter Kunz, DoctorBox, Germany; Takakazu Kurokawa, National Defence Academy, Japan; Rauno Kuusisto, Finnish Defence Force, Finland; Martti Lehto, National Defence University, Finland; Mr Trupil Limbasiya, NIIT University, Neemrana, Rajasthan, India; Dr Efstratios Livanis, University of Macedonia, Greece; Dr Leandros Maglaras, Full time academic, De Montfort University; James Malcolm, University of Hertfordshire, UK; Dr Mary Manjikian, Regent University, USA; Dr Arianit Maraj, Lecturer, AAB College-Faculty of Computer Sciences; Mario Marques Freire, University of Beira Interior, Covilhã, Portugal; Ioannis Mavridis , University of Macedonia, Greece; Rob McCusker, Teeside University, Middlesbrough, UK; Dr Imran Memon, zhejiang university, china; Dr Shahzad Memon, University of Sindh, Pakistan; Jean-Pierre Molton Michel, Ministry of Agriculture, Haiti; Dr. Yonathan Mizrachi, University of*

*Haifa, Israel; Dr Pardis Moslemzadeh Tehrani, University of Malaya, Malaysia; Evangelos Moustakas, Middlesex University, London, UK; Antonio Muñoz, University of Málaga, Spain; Daniel Ng, C-PISA/HTCIA, China; Dr. Funminiyi Olajide, Nottingham Trent University, UK; Dr Cyril Onwubiko, Cyber Security Intelligence at Research Series Limited,, UK; Rain Ottis, Tallinn University of Technology, Estonia; Dr Mahmut Ozcan, Webster University, USA; Prof Teresa Pereira, Instituto Politécnico de Viana do Castelo, Portugal; Michael Pilgermann, University of Glamorgan, UK; Dr Bernardi Pranggono, Sheffield Hallam University, UK; Prof Carlos Rabadão, Politechnic of Leiria, Portugal; Dr. Muttukrishnan Rajarajan, City University London, UK; Prof Saripalli Ramanamurthy, Pragati Engineering College, India; Dr Trishana Ramluckan, University of KwaZulu-Nata, South Africa; Dr Aunshul Rege, Temple University, United States; Dr. Neil Rowe, US Naval Postgraduate School, Monterey, USA; Prof Vitor Sa, Catholic University of Portugal, Portugal; Dr. Char Sample, Carnegie Mellon University/CERT, USA; Prof. Henrique Santos, University of Minho, Portugal; Prof Leonel Santos, Full time academic, Polytechnic of Leiria; Dr Keith Scott, De Montfort University, UK; Prof. Dr. Richard Sethmann, University of Applied Sciences Bremen, Germany; Dr. Yilun Shang, Northumbria University, UK; Prof. Paulo Simoes, University of Coimbra, Portugal; Dr Umesh Kumar Singh, Vikram University, Ujjain, India; Prof. Jill Slay, University of South Australia, Australia; Dr Lee Speakman, University of Chester, UK; Dr Joseph Spring, University of Hertfordshire, UK; Dr Hamed Taherdoost, Hamta Group, Hamta Business Corp, Vancouver, Canada; Unal Tatar, University at Albany - SUNY, USA; Dr. Selma Tekir, Izmir Institute of Technology, Turkey; Prof. Dr. Peter Trommler, Georg Simon Ohm University Nuremberg, Germany; Prof Tuna USLU, Istanbul Gedik University, Occupational Health and Safety Program, Türkiye; Craig Valli, Edith Cowan University, Australia; Dr Brett van Niekerk, University of KwaZulu-Natal & Transnet, South Africa; Richard Vaughan, General Dynamics UK Ltd, UK; Dr Namosha Veerasamy, Council for Scientific and Industrial Research, South Africa; Dr Sangapu Venkata Appaji, KKR & KSR Institute of Technology and Sciences, India; Stilianos Vidalis, School of Computer Science, University of Hertfordshire, UK; Dr. Natarajan Vijayarangan, Tata Consultancy Services Ltd, India; Dr Khan Ferdous Wahid, Airbus Group, Germany; Dr. Santoso Wibowo, Central Queensland University, Australia; Prof. Trish Williams, Flinders University, Australia; Prof Richard Wilson, Towson University, USA; Simos Xenitellis, Royal Holloway University, London, UK.*

## Biographies

### Conference and Programme Chairs



**Dr Thaddeus Eze** is a Senior Lecturer in Cyber Security at the University of Chester. He is the founder and convener of the IEEE UK & Ireland YP Postgraduate STEM Research [Symposium](#), Vice Chair, IEEE UK & Ireland Young Professionals, a technical committee member for a number of international conferences (e.g., ICAS, CYBERWORLDS, EMERGING etc.), and currently runs the Computer Science departmental [research seminar series](#). His research interests include Trustworthy Autonomics, MANET and Cyber Security (specifically, Return Oriented Programming, Policing the Cyber Threat and Cyber Education) and he has a number of publications in these areas.



**Dr Lee Speakman** is a Senior Lecturer and Programme Lead for Cybersecurity at the University of Chester. Lee has worked in Defence since 1999. He has gained experience in Intelligence, Strategy, Electronic Warfare, Communications, and Networking. He gained his PhD in MANETs from Niigata University, Japan, in 2009, and returned to Defence to work in Cyber and Information Security, and related areas. He joined the University of Chester in 2015 as the Programme Leader to develop and deliver Cybersecurity courses. His research interests are in software and system security.



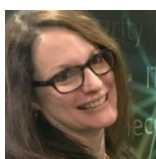
**Dr Cyril Onwubiko** is the Secretary, IEEE UK & Ireland, Chair, IEEE UK & Ireland Blockchain Group, and Director, Cyber Security Intelligence at Research Series Limited, where he is responsible for directing strategy, IA governance and cyber security. Prior to Research Series, he had worked in the Financial Services, Telecommunication, Health sector and Government and Public services Sectors. He is a leading scholar in Cyber Situational Awareness (Cyber SA), Cyber Security, Security Information and Event Management (SIEM), Data Fusion & SOC; and interests in Blockchain and Machine Learning. He is the founder of the Centre for Multidisciplinary Research, Innovation & Collaboration (C-MRiC) <https://www.c-mric.com>. Detailed profile for Cyril can be found on <https://www.c-mric.com/cyril>

### Keynote Speakers

**Phil Byrom** is a Merseyside Police Officer with 19 years' service. Phil is currently seconded to the North West Regional Organised Crime Unit as a Cyber Crime Prevent Officer and is responsible for delivering cybercrime prevention work across 6 North West Police forces. He also works to identify individuals who are on the verge of moving into cybercrime or have committed low level cybercrime and prevent them from either moving into cybercrime or divert them onto positive pathways and use their computer skills in more positive ways.



**Will Farrell** is a Police Detective with 12 years' experience, currently working with Cyber Crime Prevent at the North West Regional Organised Crime Unit in the UK. Will is responsible for delivery cybercrime prevention work across the 6 police forces in North West England. Will works to identify individuals on the cusp of committing cybercrime, and once identified, Will seeks to deter and divert them towards more positive pathways. Before joining the police, Will spent fourteen years in the commercial radio industry, managing the programme output of five radio stations. He was also a successful radio DJ for many popular radio stations.



**Louisa Murphy** is a Regional Cyber Protect Officer in the Cyber Crime Unit at the North West Regional Organised Crime Unit. (A collaboration of 6 local police forces). As part of a national policing network, her role focusses on making the North West a safer place online. She links directly to both the National Cyber Security Centre (NCSC) and Action Fraud to obtain the latest cybercrime information and will help organisations to stay on top of the latest cyber security information. Louisa works with the different sectors promoting and encouraging cyber resilience within their organisations. She speaks at conferences and business events to raise awareness of current cyber threats and national advice and best practice so that businesses and individuals can best protect themselves and their



information and assets. She also works with victims to understand how they were targeted and to help them become more cyber secure.



**Keith Terrill** is a Regional Cyber Protect Officer in the North West Regional Cyber Crime Unit. They're responsible for delivering the Protect strand of UK Cyber Policing in the North West in conjunction with local forces, Action Fraud and the National Cyber Security Centre. The goal of this is to raise awareness of the latest cyber crime threats impacting businesses and individuals, and highlighting the many practical steps that can be taken to improve cyber resilience. Outside of his role in the NWRCCU Keith has served as a Special Constable (volunteer Police Officer) in Response and Neighbourhood policing for nearly three years and previously worked as Programmer specialising in Networking.

**Detective Inspector David Turner** has over twenty years of experience in investigating serious, major and organised crime including counter terrorism. David Turner is presently responsible for managing the North West Regional Cyber Crime Unit to deliver against the Serious Organised Crime Strategy 2018 and National Cyber Security Strategy 2016 – 2021. Their role involves managing all the staff within the Regional Cyber Crime Team including the Regional Coordinator, the Regional Crime Cyber Unit 4P capability as well as the Dark Web and Digital Forensics Team. They have oversight of all investigations in the RCCU along with Protect, Prevent and Prepare activity ensuring collaboration across a range of sectors from law enforcement overseas to third sector organisations in the UK.

### Mini Track Chairs



**Dr. Sangapu Venkata Appaji** is working as a Professor, at KKR and KSR Institute of Technology and Sciences, Guntur, India. He completed his doctorate in Computer science and Engineering in the area of Cryptography and Network Security from Jawaharlal Nehru Technological University Hyderabad, INDIA. He has more than 10 years of teaching experience in Computer Science and engineering and Information Technology department for graduate and post graduate students. He supervised IOT, Security related projects for graduate and postgraduate students.



**Dr. Shahzad Memon** is working as a Professor, at department of Electronics, Faculty of Engineering and Technology at University of Sindh, Pakistan. He completed his Doctorate PhD in Biometrics security from Brunel University, London, UK. He published his research in several national and international research journals. Dr. Memon attended and presented his research in national and international conferences organized in USA, UK and Europe. Dr. Memon supervised MS and PhD students in the field of Biometrics, Cyber Security and Privacy, Smarts systems security and Cyber Physical systems security. He also granted funding for research from Higher Education commission and ICT Research and Development, Ministry of Information Technology, Pakistan..



**Dr Brett van Niekerk** is a senior lecturer in computer science at the University of KwaZulu-Natal. He serves as chair for the International Federation of Information Processing Working Group on ICT in Peace and War, and the co-Editor-in-Chief of the International Journal of Cyber Warfare and Terrorism. He has numerous years of information/cyber-security experience in both academia and industry, and has contributed to the ISO/IEC information security standards. In 2012 he graduated with his PhD focusing on information operations and critical infrastructure protection. He is also holds a MSC in electronic engineering and is CISM certified.



**Dr Trishana Ramluckan** is a Postdoctoral Researcher in the School of Law and an Adjunct Lecturer in the Graduate School of Business at the University of KwaZulu-Natal. She is a member of the IFIP working group on ICT Uses in Peace and War, the Institute of Information Technology Professionals South Africa and is an Academic Advocate for ISACA. In 2017 she graduated with a Doctor of Administration specialising in IT and Public Governance and in 2020 she was listed as in the Top 50 Women in Cybersecurity in Africa. Her current research areas include Cyber Law and Information Technology Governance.

## Workshop Facilitator



**Dr Edwin “Leigh” Armistead** is the President of Peregrine Technical Solutions, a certified 8(a) small business that specializes in Cyber Security. A retired United States Naval Officer, he has significant Information Operations academic credentials having written his PhD on the conduct of Cyber Warfare by the federal government and has published three books, in an unclassified format in 2004, 2007 and 2010, all focusing on full Information Warfare. He is also the Chief Editor of the Journal of Information Warfare (JIW) <https://www.jinfowar.com/>; the Program Director of the International Conference of Cyber Warfare and Security and the Vice-Chair Working Group 9.10, ICT Uses in Peace and War. Shown below are the books on full spectrum cyber warfare and the JIW

## Biographies of Contributing Authors

**Kari Alenius** is Professor in General History and Head of Department at the University of Oulu (Finland). He also worked as a Visiting Professor at Lakehead University, Canada, in 2019–2020. He has specialized on the history of Eastern Europe, history of ethnic relations, and history of information warfare.

**Mark Baggett** Vice President, Industrial Control Systems (ICS), Mission Secure, Mark’s an industry veteran and ICS expert. He’s designed, engineered, and implemented control systems internationally for energy’s most prominent players. Mark leverages his expertise to help operations assess and mitigate cyber risks and implement a secure architecture, managing OT cybersecurity projects for rigs, refineries, pipelines, manufacturing plants, and chemical facilities.

**Johnny Bengtsson** is a forensic expert in hardware forensics at Swedish National Forensic Centre (NFC), and part-time industrial PhD student in IoT forensics at Linköping University, Sweden. He holds a Master of Science degree in Electrical Engineering from Linköping University, and also a University Diploma in Chemical Engineering from Chalmers University of Technology, Sweden.

**George-Daniel Bobric** is a PhD candidate within the “Carol I” National Defence University, Romania. His main areas of interest are mathematical and computer sciences, cyber security and information warfare.

**Micki Boland** is a global cybersecurity warrior and evangelist with Check Point Software Technologies Office of the CTO. A practitioner with 20 years in ICT, cybersecurity, emerging technology innovation, Micki holds ISC2 CISSP, Master of Science in Technology Commercialization from the University of Texas at Austin, MBA with Global Security concentration from East Carolina University.

**Long Chen** is currently pursuing the Ph.D. degree with the Beihang University, under the supervision of Prof. C. Xia. He has participated in several National Natural Science Foundations and other research projects as a Director and Contributor. His research interests include Network and Information Security, Intrusion Detection Technology.

**Dr. Jim Q. Chen**, Ph.D: Professor of Cyber Studies, College of Information and Cyberspace (U.S. National Defense University) Expertise in cyber warfare, cyber deterrence, cyber strategy, cybersecurity technology, artificial intelligence, and machine learning. Authored and published numerous peer-reviewed papers, articles, and book chapters on these topics. Has also been teaching graduate courses on these topics. A recognized expert in cyber studies and artificial intelligence.

**Dr. Sabarathinam Chockalingam** is a Research Scientist at the Institute for Energy Technology in Halden, Norway. Saba has a PhD in Cyber Security from the Delft University of Technology and MSc in Cyber Security and Management from the University of Warwick. His research interests include cyber security, risk management and serious games.

**Joel Coffman** is an Associate Professor in the Department of Computer and Cyber Sciences at the US Air Force Academy. He received his BS in computer science from Furman University, and his MS and PhD in computer science from the University of Virginia. Joel’s research interests include automated software diversity, cloud security, and keyword search in databases.

2d Lieutenant **Tristan Creek** is completing his Masters degree in Cyber Operations at the Air Force Institute of Technology, Wright-Patterson AFB, OH, USA. Research interests include covert communication channels, long range WiFi exploitation, and Bluetooth Low Energy exploitation.

**Tiago Cruz**: Ph.D. degree in informatics engineering (University of Coimbra, 2012), where he has been an Auxiliary Professor with the Department of Informatics Engineering, since 2013. Research interests cover management systems for communications infrastructures and services, critical infrastructure security, broadband access network device and service management, Internet of Things, software defined networking, and network function virtualization.

**Theodore Drewes** is a senior at the United States Air Force Academy with a major in computer science. He plans on graduating in May 2021 and continue his career as a RPA pilot. While at the Academy, he took an interest in Artificial Intelligence, malware development, and computer graphics.

**Jaco Du Toit** is a lecturer at the Academy of Computer Science and Software Engineering at the University of Johannesburg. His areas of research include Cyber Security, with a focus on privacy and mobile operating environments. A specific interest to him is research in increasing the protection of private information using decentralised data and access control models.

**Dr. Jan Fesl** obtained his Ph.D. diploma in 2018 at Czech Technical University in Prague. His areas of research are computer networks, distributed systems, cyber-security. Dr. Fesl is the leader of the Networking research group from the Faculty of Information Technology at Czech Technical University in Prague.

**Adrian Ford** is an information technology manager with over 25 years' experience and a doctoral research student of information security at the University of East London. He holds an MBA from Lancaster University Management School (2009), professional membership of the British Computer Society (MBCS) and is a Freeman of the Worshipful Company of Information Technologists.

**Tim Grant** is retired but an active researcher (Professor emeritus, Netherlands Defence Academy). Tim has a BSc in Aeronautical Engineering (Bristol University), a Masters-level Defence Fellowship (Brunel University), and a PhD in Artificial Intelligence (Maastricht University). Tim's research focuses on offensive cyber operations and on Command & Control and Emergency Management systems. More details can be found at <https://www.linkedin.com/in/tim-grant-r-bar/>.

**Dr Babak Habibnia** PhD in Computer Science with a specialization in Digital Forensics and Cybercrime Investigation. His academic research focuses on redesigning digital forensics as a computer-assisted human activity. Dr Babak Habibnia presents for the first time a new scientific semi-automated approach based on visualization of relevant data properties to help investigators detect digital evidence tampering and anomaly.

**Dr. Aarne Hummelholm**, PhD in Information Technology (University of Jyväskylä, 2019). He has over 30 years' experience in the design, development of architectures' of authorities' telecommunications networks and information systems. Key themes in his work have been critical service availability, usability, cyber security and preparedness issues.

**Angela G. Jackson-Summers** is an Assistant Professor of Information Systems in the Management Department at the U. S. Coast Guard Academy. She received her Ph.D. in Business Administration (Information Systems) from Kennesaw State University. Her research interests include IT/IS risk management, and data/information security and assurance.

**Dr Victor J Jaquire** has been within the field of cyber security for over 20 years within Government and the Private sector. He holds an Honours Degree in Management from Henley University and a Master's and PhD in Informatics from the University of Johannesburg. He has published various academic papers on cyber strategies and cyber counterintelligence maturity.

**Andrew Johnson** is a 3rd year PhD student within the Cyber Security Department of the University of South Wales, UK. His research field is predominantly in the modelling of Side Channel Attacks. Andrew continued his

research study at the University after completing his MSc in Computer Systems Security in 2018. He has previously published two conference papers with the IEEE during his PhD study.

**Thorsten Kodalle** LTC (General Staff) lectures on security policy at the Command and Staff College of the German Armed Forces with a particular focus on NATO, Critical Infrastructure and Cyber. He is a member of the NATO research task group “Gamification of Cyber Defense/Resilience”, an experienced facilitator of manual wargaming on the operational level for courses of action analysis, for operational analysis, operations research, serious gaming and especially for matrix wargaming.”

**Eleni Kapsokoli** is a Ph.D. candidate at the University of Piraeus, Department of International and European Studies, Greece and Ph.D. Fellow of the European Doctoral School. She holds a bachelor's degree in Political Science and Public Administration and a master's degree in International Relations and Strategic Studies. Her main research interests are international security, terrorism, cybersecurity and cyberterrorism.

**Nickson M. Karie** received his PhD degree in computer science from the University of Pretoria, South Africa. Currently, he is a cybersecurity research fellow at Edith Cowan University, Perth, Australia. He has over 10 years of experience in academic teaching, research, and consultancy. His research interests include network security and forensics, intrusion detection and prevention, cloud and IoT security

**Joakim Kävrestad** is a doctoral student, at the University of Skövde, focusing on human aspects of cybersecurity. He is a prior forensic expert who is coordinating a master's program in Privacy, Information and Cybersecurity and teaching classes in digital forensics and technical cybersecurity.

**Jan Kleiner** is a PhD student of Political Science at Masaryk University in the Czech Republic. He focuses primarily on cybersecurity, the relationship between a state and citizens in cyberspace (e.g., how states secure their citizens in cyberspace), and propaganda and information warfare. He mainly employs quantitative (statistical) and mixed methods research designs.

**MSc. Tiina Kovanen**, MSc Tiina Kovanen is a PhD student at the university of Jyväskylä. She is interested in various cyber security topics for different cyber-physical systems. Currently she is working towards her degree by studying possibilities and challenges related to ships remote pilotage environment, ePilotage.

**Dr. Sylvain (Sly) Leblanc** is a Professor in Computer Engineering at the Royal Military College of Canada, also serving as Chair for Cybersecurity and Primary Investigator of the Computer Security Laboratory. His research interests are in the Cyber Security of Vehicular Systems, Network Counter-Surveillance Operations, Vulnerability & Security Assessments and Cyber Education.

**Louise Leenen's** areas of specialisation are Artificial Intelligence applications in cybersecurity and mathematical modelling. She is currently an Associate Professor in the Computer Science Department at the University of the Western Cape in South Africa and a member of the Centre for Artificial Intelligence (CAIR).

**Faith Lekota** is an Independent IT Consultant. She is a PhD candidate at the University of Johannesburg. Her research interest include cybersecurity frameworks, and information security best practise standards.

**Dr. Andrew N. Liaropoulos** is Assistant Professor in University of Piraeus, Department of International and European Studies, Greece. He is also a senior analyst in the Research Institute for European and American Studies (RIEAS) and a member of the editorial board of the Journal of European and American Intelligence Studies (JEAIS).

**Capt Banks Lin, USAF** (BS, San Jose State University) previously served as cybersecurity test lead at the Air Force Operational Test & Evaluation Center, conducting operational-realistic cyber assessments on space and missile weapon systems. He is currently a student at Air Force Institute of Technology studying for a Master of Science in Cyber Operations.

**Christoph Lipps, M.Sc.** graduated in Electrical and Computer Engineering at the University of Kaiserslautern where he meanwhile lectures as well. He is a Researcher and Ph.D. candidate at the German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern. His research focuses on Physical Layer Security (PhySec),

Physically Unclonable Functions (PUFs), Artificial Intelligence (AI), entity authentication and all aspects of network and cyber security.

**Dr. Arianit Maraj**, is a professor of Engineering Informatics and is also in Telecom of Kosovo. He received PhD from Polytechnic University of Tirana, in 2013. His research interest lay in Data security, Wireless communications and Ad-Hoc networking. He has published a considerable number of scientific papers on international journals and conferences.

**Erik David Martin** BSc is currently working as an Infrastructure and Security Engineer at Sopra Steria in Stavanger, Norway. He has submitted several entries in the Exploit Database and has been active in the security research community throughout the last years. He has also published a series of articles regarding IoT and SCADA security. His research interests lie in the area of computer security and computer forensics.

**Dr. Edward Mienie** is the Executive Director of the Strategic & Security Studies program at the University of North Georgia. In addition, as associate professor he teaches national intelligence courses within his degree program. He has most recently helped introduce national intelligence education as an elective course to Georgia high schools, one of the first in the nation.

**Jitesh Miglani** B. Tech, M.Sc. graduated with a B.Tech in Computer Science Engineering from The NorthCap University, India in 2017 and a MSc in Applied Cyber Security from Technological University Dublin. He is currently working as a cyber security consultant in Deloitte Ireland.

**Masombuka Mmalerato** is cybersecurity specialist and currently working on her PhD with the main focus on Artificial Intelligence and cybersecurity. She has co-authored and peer-reviewed several articles on these disciplines. Her other research interests include blockchain, quantum computing and data analytics.

**Haya Yusuf Mohamed** Gulf Air Company, and MBA Student. College of Business and Finance, Ahlia University, Manama, Bahrain

**Fredrick Ochieng' Omogah** is a lecturer of I.T & Medical Informatics at Uzima University, Kenya. He is currently finalising Msc. I.T Security and Audit from Jaramogi Oginga Odinga University, Kenya. Received Bachelor of I.T from Australia, 2009. His main research areas are in I.T and cyber security in electronic healthcare

**Dr Olav Opedal** is an independent psychologist and data scientist in Ellensburg, WA, US. He received his PhD in General Psychology from Capella University in 2019. His main research areas are personality and behavior associated with computer use, and the applied use of big data analytics, machine learning and AI as an independent ML/AI practitioner.

**Jani Päijänen** (B.Eng) works as a Project Manager at the Institute of Information Technology of JAMK University of Applied Sciences. He has experience from delivering consultancy for clients in Project Management, Information Technology, and Software Development. Jani is currently doing his M.Eng in Cyber Security.

**Dr. Bryson Payne** is a nationally-recognized cyber coach, author, TEDx speaker, and the founding Director of the Center for Cyber Operations Education at the University of North Georgia, an NSA-DHS Center for Academic Excellence in Cyber Defense. He has coached programming and cyber competition teams at UNG since 2005, including UNG's #1 in the nation NSA Codebreaker Challenge 2019 and 2020 cyber operations teams.

**Alexander Pfeiffer**: recipient of a Max Kade Fellowship awarded by the Austrian Academy of Science to work at the Massachusetts Institute of Technology (MIT), Department for Comparative Media Studies / Writing in 2019 and 2020. In 2021 he returned to Donau-Universität Krems. Currently approaching his second PhD at the department of AI at the University of Malta. <https://www.alexpfeiffer.at>

**Pranav Phadke**, B.Sc, M.Sc. Pranav Phadke graduated with a B.Sc. in Information Technology and Masters In Computer Application from the University of Mumbai in 2015 and an M.Sc in Applied Cyber Security in 2019 from the Technological University of Dublin. He is currently working as a Software Engineer with a firm based in Dublin.

**Captain Carl Poole** received a master's degree in space systems with specialties in space vehicle design and space control modelling and simulation from the Air Force Institute of Technology March 2021. His research topics include the examination of space-based ballistic missile defense architectures for employed kinetic weapon concepts.

**Lucas Potter** is a Biomedical Engineering PhD Student and member of the SAMPE (Systems Analysis of Metabolic Physiology) Lab at Old Dominion University. His doctoral research is focused on cellular respiration in those with compromised metabolism. Past research endeavors include human factors research, specifically human factors analysis of performance in virtual reality, modeling of physiology, and materials engineering.

**Seyedali Pourmoafi** I have spent most of my time learning and build up my knowledge around Computer Science subject ever since I lay hands on my first self-build computer at my early childhood self-studying desire and researching on the internet led me to learn very fast in childhood. I am extremely curious about Astronomy and Computer science. I decided to study Mathematics in high school and Computer Science at the university. I had my first degree in Computer Science software engineering before I moved to the UK. Then I decided to start with an Undergraduate degree and choosing an entirely new degree and specialism which leads to being graduated in information technology with a Web specialism degree. Then after having a short discussion with a member of the faculty, I decided to follow the new direction by choosing M.Sc. in networking, now I am finishing up my Ph.D. study in Cybersecurity.

**Jakub Pražák** is a Ph.D. candidate of International Relations at the Charles University's Faculty of Social Studies and a project assistant at the Prague Security Studies Institute. His main research areas are weaponization of outer space and dual-use technology.

**Carlos Rabadão** is Coordinator Professor at Polytechnic Institute of Leiria. He received his PhD degree in Computer Engineering from University of Coimbra in 2007. He has published more than 50 papers at conference proceedings and refereed journals. His major research interests include Cybersecurity, Information Security Management Systems and Intrusion Detection Systems for IoT.

**Jyri Rajamäki** is Principal Lecturer in Information Technology at Laurea University of Applied Sciences and Adjunct Professor of Critical Infrastructure Protection and Cyber Security at the University of Jyväskylä, Finland. He holds D.Sc. degrees in electrical and communications engineering from Helsinki University of Technology, and a PhD in mathematical information technology from University of Jyväskylä.

**Dr Trishana Ramluckan** is the group research manager for Educor Holdings. In 2020 she completed post-doctoral research in International Cyber Law at the School of Law, UKZN. Her areas of research include the intersections of IT with law and governance. She is a member of the IFIP working group on ICT Uses in Peace and War and is an Academic Advocate for ISACA.

**Dr Harri Ruoslahti** is a Senior lecturer of Security and risk management at Laurea University of Applied Sciences, a researcher in related projects, and Laurea's point of contact in ECHO (the European network of Cybersecurity centres and competence Hub for innovation and Operations).

**Karo Saharinen** (M.Eng) is working as a Senior Lecturer in IT and handling the responsibility of degree programme coordinator of the master's degree programme in IT, Cyber Security at JAMK University of Applied Sciences. He is currently working on his PhD related to Cyber Security Education.

**Leonel Santos** is Assistant Professor at Polytechnic Institute of Leiria. He received his PhD degree in Computer Science from University of Trás-os-Montes e Alto Douro in 2020 and is a researcher at Computer Science and Communication Research Centre. His major research interests include Cybersecurity, Information and Networks Security, IoT, Intrusion Detection Systems and Computer Forensics.

**Janine Schmoldt** studied International Relations at the University of Erfurt, Germany. Afterwards, she completed her Master at the Vrije Universiteit Amsterdam where she studied Law and Politics of International Security. She is currently a PhD student at the University of Erfurt.

**DR KEITH SCOTT** is Programme Leader for English Language at De Montfort University in Leicester. His research operates at the intersection of communication, culture and cyber, with particular interests in influence, information warfare, and simulations and serious gaming as a training, teaching, and research tool.

**Jussi Simola** is a doctoral candidate of cyber security at University of Jyväskylä. He received his master degree in Information Systems from the Laurea University of Applied Sciences in 2015. He has worked as a cybersecurity specialist and he has participated in the development of a common Early Warning System for the EU member countries.

**Dr. Joshua Alton Sipper** is a Professor of Cyberwarfare Studies at the Air Force Cyber College. He completed his Doctoral work at Trident University in September of 2012, earning a Ph.D. in Educational Leadership (emphasis, E-Learning Leadership). Dr. Sipper's research interests include cyber ISR, policy, strategy, and warfare.

**Abderrahmane Sokri** has a Ph.D. in administration from HEC-Montreal. He is currently serving as economist for the Canadian Department of National Defence. His current research interest includes game theory applied to military operations. He has published in good international journals such as the European Journal of Operational Research.

**Sanjana Suresh** is a freshman at the LeBow College of Business within Drexel University, majoring in Finance and Business Analytics. She is actively involved in a variety of activities at Drexel, including Drexel's Undergraduate Student Government Association, Drexel Women in Business, and Undergraduate Research. In her free time, Sanjana enjoys spending time with her friends and family, reading, writing, and playing lacrosse.

**Professor Iain Sutherland** BSc MSc PhD MBCS is currently Professor of Digital Forensics at Noroff University College in Kristiansand, Norway. He is a recognised expert in the area of computer forensics. He has authored articles ranging from forensics practice and procedure to network security. His current research interests lie in the areas of computer forensics and computer security.

**Christina Thorpe**, B.Sc, Ph.D. Christina Thorpe graduated with a B.Sc. (Hons) in Computer Science from University College Dublin in 2005 and a Ph.D. in Computer Science in 2011. She was a postdoctoral research fellow in the Performance Engineering Lab in UCD from 2011 - 2018. She is currently a Lecturer in Cyber Security in the Technological University Dublin.

**Mr. Ilkka Tikanmäki** is a researcher at Laurea University of Applied Sciences and a doctoral student of Operational art and tactics at the Finnish National Defence University. He holds an MBA degree in Information Systems and BSc degree in Information Technology.

**Maija Turunen** is a postgraduate student in military sciences at the Finnish National Defense University. Her main research areas consist of cyber warfare and Russia. Maija Turunen works as a legal counsel at the Finnish Transport Infrastructure Agency.

**Brett van Niekerk** is a senior lecturer at the University of KwaZulu-Natal. He serves as chair for the IFIP Working Group on ICT in Peace and War, and co-Editor-in-Chief of the International Journal of Cyber Warfare and Terrorism. He holds a PhD focusing on information operations and critical infrastructure protection.

**Namosha Veerasamy**: BSc: IT Computer Science Degree, BSc: Computer Science (Honours Degree), MSc: Computer Science with distinction (University of Pretoria) and a PhD (University of Johannesburg). Currently senior researcher (Council for Scientific and Industrial Research in, Pretoria). Qualified as a Certified Information System Security Professional and Certified Information Security Manager. Has been involved in cyber security research and governance for over 15 years.

**Gábor Visky** is a researcher at NATO CCDCOE, his main field of expertise is industrial control systems. Gábor's previous assignments include 15 years of designing hardware and software for embedded control systems and researching their vulnerabilities through reverse engineering. Gábor holds an MSc degree in Information Engineering with a specialty in Industrial Measurement.

**Samuel Wairimu** is a PhD student in Computer Science in the Department of Mathematics and Computer Science at Karlstad University, Sweden. He received his Master's in Cybersecurity from the University of Chester, UK in the year 2018. His main research areas are cybersecurity, cyberwarfare, information security and privacy, and security and privacy in e-Health.

**Richard L. Wilson** is a Professor of Philosophy at Towson University in Towson, MD. Teaching Ethics in the Philosophy and Computer and Information Sciences departments and Senior Research Fellow in the Hoffberger Center for Professional Ethics at the University of Baltimore. Professor Wilson specializes in Applied Ethics teaching a wide variety of Applied Ethics Classes.

**Ashley Wood** is a Visiting Lecturer and current PhD student at the University of Chester, with a First class BSc(Hons) degree in Cybersecurity and an MSc in Advanced Computer Science with Distinction. Ashley has keen interests in digital forensics, cybercrime investigation, systems/network security and malware analysis.



# The PUF Commitment: Evaluating the Stability of SRAM-Cells

Pascal Ahr<sup>1</sup>, Christoph Lipps<sup>1</sup> and Hans Dieter Schotten<sup>1,2</sup>

<sup>1</sup>German Research Center for Artificial Intelligence, Kaiserslautern, Germany

<sup>2</sup>University of Kaiserslautern, Division of Wireless Communication and Radio Positioning, Kaiserslautern, Germany

[Pasacal.Ahr@dfki.de](mailto:Pasacal.Ahr@dfki.de)

[Christoph.Lipps@dfki.de](mailto:Christoph.Lipps@dfki.de)

[Hans\\_Dieter.Schotten@dfki.de](mailto:Hans_Dieter.Schotten@dfki.de)

DOI: 10.34190/ECW.21.031

**Abstract:** Static Random Access Memory (SRAM) based Physical Unclonable Functions (PUFs) are a dedicated sub-area of silicon PUFs in the research area of Physical Layer Security (PhySec). Due to their high Shannon Entropy, low energy consumption and availability they are particularly suitable for Industrial Internet of Things (IIoT) security applications. SRAMs are volatile memories, bistable systems which always adopt one of two values: zero or one. During the startup process - powering up the cells-, the cells take one of these states, the so called Startup-Value. This “hardware fingerprint” is depending due to physical features, fluctuations and deviation occurring during the manufacturing process of the semiconductors and the devices, and can thus be different at each restart. For a function in a mathematical meaning, and particularly for cryptographic applications, it is necessary that every element of the definition area is only mapped to one element of the codomain. For this purpose the startup-values of the SRAM have to be (mostly) stable for every restart. To verify the suitability, and appropriateness for cryptographic applications, the paper examines the stability of the startup-values; how often does the same but still individual bit-patterns occur and how many and which bits are flipping. To provide comparable results, 30 SRAMs are evaluated with 500 startup procedures each. For automated testing a Printed Circuit Board (PCB) is implemented, controlled by a Microcontroller Unit (MCU). In order to monitor the temperature and humidity –as external influencing factors of the startup behaviour- corresponding sensors are integrated as well. The evaluation provides a high resolution of the course of stability over the various measures, and thus enables a detailed analysis. As a part, the mapping of functions to data-points is done by using regression tools. Thereby it is not only possible to determine the stability in total, but the course over all restarts as well. The results of the work contribute to PUF research in general and prove the applicability of SRAM-PUFs in IIoT and other application areas, especially for resource constrained devices, by evaluating and proving the stability of SRAM cells.

**Keywords:** physical layer security, physically unclonable functions, SRAM-stability, industrial internet of things, cyber security

## 1. Authentication in Industrial Internet of Things Environments and the Concept of Physical Layer Security

Due to the changing requirements in industrial manufacturing environments, towards Industrial Automation and Control Systems (IACSs), Cyber-Physical Production Systems (CPPS) and the Industrial Internet of Things (IIoT), the security requirements of the systems are changing as well. New application scenarios such as Machine-to-Machine (M2M) and Machine-to-Service (M2S) communication arise, driven by the ever increasing amount of inter-connected devices (Farooq, et al., 2015). However, the fundamental need for secure, reliable and available systems persists. On the contrary, the supplementary requirements of the industry –available bandwidth, real time processing and the limited resources of the devices- further complexify the task. Besides that, in industrial use-cases the live cycles are much longer compared with Commercial-of-the-Shelf (COTS) hardware components, whereby it is crucial task to establish and integrate security mechanism into existing systems.

As the overall necessity to abolish and replace (traditional) passwords is requested by (Bonneau, et al., 2012), a worthwhile approach to meet these requirements consists of the Physical Layer Security (PhySec) concept. This is in particular a set of different algorithms based on the same operating principle of utilizing existing physical properties to derive cryptographic primitives. At a high level of abstraction, three concepts are distinguished: Human – Physically Unclonable Functions (H-PUFs) (Lipps, Herbst & Schotten, 2021), Channel-PUFs (Lipps, et al., 2020) and Electric or Silicon PUFs (Gassend, et al., 2002). The Static Random-Access Memory (SRAM)-based PUFs, addressed in this work, are one subset of the Silicon PUFs. Compared to conventional security approaches such as Trusted Platform Modules (TPMs) (Hoeller & Toegl, 2018), certificates (Lee, Alasaarela & Lee, 2014) and asymmetric cryptography (Schneier, 2015), they offer benefits such as low acquisition costs –SRAMs are already built into almost every electronic device-, the time it takes for the PUF to be available on the device –a few nanoseconds- and the variable usability. Summarized, the approach of using SRAMs is the resource efficiency, as

no complex computations are required to calculate cryptographic credentials, the “secrets” are already given by the circuit. In addition, the keys do not need to be stored in Non-Volatile Memory (NVM) –key storage is also vulnerable to attacks and prone to reengineering– as the keys can be regenerated every time required.

Since the general introduction of the PUF concept by Gassend et al. (2002) meanwhile a lot of derivatives exist. These include Ring-Oscillator PUFs (Gao & Qu, 2014), Field Programmable Gate Array (FPGA) intrinsic PUFs (Guajardo, et al., 2007) and Optial-PUFs (Rührmair, et al., 2013). As further discussed in Section 2, slight deviations and inaccuracies during the doping of semiconductor devices (Halak, 2018) add up to a chip individual behavior: the startup behavior of the SRAM cell. The strength results from the unpredictability of this behaviour, as the deviations are so small that they cannot be influenced and mimicked (Kusters & Willems, 2019), which is why these structures are called “an objects fingerprint” (Maes, 2013) and are defined as “a physical entity whose behaviour is a function of its structure” (Halak, 2018).

Ning et al. (2019) give an overview of PUFs in general, as well as the different structures and architectures, but also describe the requirements for a secure use. There, a detailed overview of the distinction between *silicon-PUFs* and *non-silicon-PUFs*, as well as the individual subgroups, is given too. Besides that a comprehensive introduction to the topic is provided by (Maes, 2013) and (Böhm & Hofer, 2013). They cover several topics of concept, introducing arbiter-, ring-oscillator and SRAM-PUFs and describe these in a mathematical manner. They conduct comprehensive evaluations of PUFs, but also provide explicit applications like authentication, identification and key generation. Helper data schemes, in particular fuzzy commitment-/ and syndrome-based schemes, for SRAM-PUFs in multi enrollment scenarios are discussed by Kusters et al. (2017). Gao et al. (2019) are rather addressing the applicability of SRAM PUFs in resource constrained devices as lightweight but secure and reliable key generation mechanism. They furthermore examine and validate fuzzy extractors and reverse fuzzy extractors while using PUFs as a seed for cryptographic hash functions and Bose-Chaudhuri-Hocquenghem (BCH) Codes. An efficient implementation of SRAM-based True Random Number Generators (TRNGs) is suggested by van der Leest et al. (2012). They use the PUF sequence as true random seed for their Deterministic Random Bit Generators (DRBGs) and evaluate their implementation by various test metrics such as the Hamming Distance test, the Context-Tree Weighting (CTW) methods test and the standard test provided by the National Institute of Standards and Technology (NIST). (Alheyasat, et al., 2020) are also addressing TRNGs and therefore define a metric in order to classify SRAM cells and select the most unpredictable cells. A hybrid cryptosystem based on the SRAM-PUF concept by deriving binary PUF vectors is given by (Lipps, et al., 2020). In combination with a Master-Slave setup a symmetric cryptosystem remains, including all benefits of Symmetric Key Cryptography (SKC). A security scheme based on the SRAM randomness and supported by appropriate error correction by polar codes is proposed by (Chen, et al., 2017), whereby they achieve a failure probability of  $10^{-9}$ . The aging of semiconductor devices based on long-term evaluation of Arduino Leonardo boards was examined by Wang, et al. (2020). The temperature behaviour of the helper data was considered by Kusters, Rikos & Willems (2020). Therefore they evaluate the key reconstruction error probability through Monte Carlo simulations. To further increase the security of conventional PUF implementations and to prevent (virtual) cloning, Prada-Delgado & Baturone (2021) propose to harden the PUF structures by additional behaviour. They validated their proposal by the combination of a physical and a behavioural response to given challenges. This approach they describe as Behavioural and Physical Unclonable Functions (BPUFs).

Based on the work of Böhm, Hofer & Pribyl (2011) about SRAM-PUFs in general and the study provided by Lipps et al. (2018) and to confirm the general usability of SRAM-PUFs in cryptographic applications this work examines the stability and the entropy of this PUF. When used in cryptographic applications, whether as a seed for TRNGs or for authentication, the resulting sequences must always be the same. This can be indicated by the stability of the cells. Moreover, the general usability is determined on the basis of the disorder in the system -the Shannon entropy. Using a statistically significant amount of SRAM devices, the stability of the cells is examined using a specially designed evaluation platform.

The remainder of this work is organized as follows. In Section 2 the used properties of the SRAM are explained, what is meant by the *Startup behaviour*, before Section 3 discusses the setup of the evaluation platform. In order to evaluate the usability, metrics such as the correlation are defined in Section 4 against which the SRAMs are analysed. The results of this evaluation are presented in Section 5 –the stability trend of the SRAM cells- and Section 6 –the distribution of the stability-, after which Section 7 concludes the work and provides an outlook to future work.

## 2. The structure of a SRAM cell and its behaviour at Startup

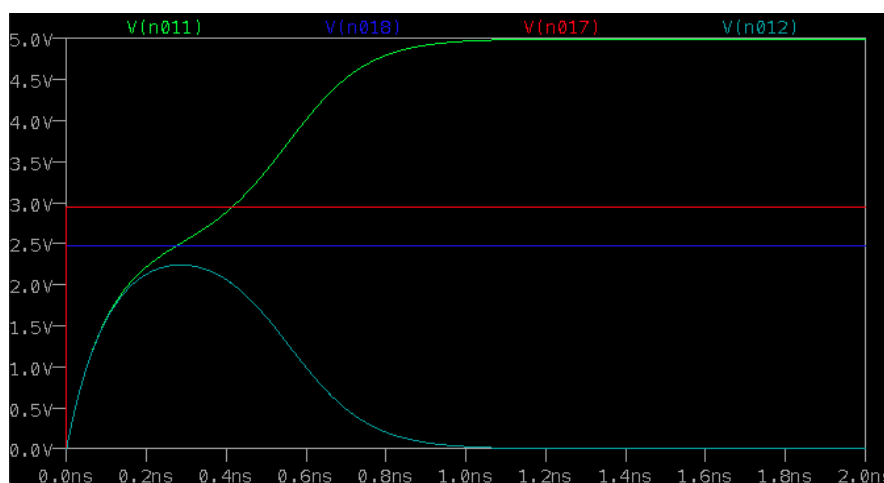
In general a SRAM cell consists mainly of two circle-coupled inverters, produced in Complementary Metal Oxid Semiconductor (CMOS) technology. Because of the coupling the cell is a bistable system, which always adopts one out of the two states –one or zero. Point  $V_{inversion}$  in **Figure 2**, is the intercept of the two transfer curves of the Inverters and divides the cell into two areas, where it adopts one of the states. If both inverters are exactly the same the intercept is on the bisecting angle and both areas are of the same size. Due to the bistability principle the cell keeps their state and therefore saves the value. To store an explicit value in the SRAM cell, a voltage must be forced at the input, which is higher or lower than the Point  $V_{inversion}$ .

The behavior of inverter's transfer curve depends on transistor parameters, among others, like the channel length, channel width, and thickness of the oxide-layer. The manufacturing of semiconductor transistors is a complex process, influenced by random technical and physical fluctuations. Due to the small size of the semiconductor structures, these fluctuations affect the parameters and causes that not all of the transistors are exactly the same. This diversity of transistors modifies the transfer curve of each inverter in that way, that two out of one SRAM cells do not have exactly the same behavior. In consequence the point  $V_{inversion}$  shifts away from the bisecting angle and the state areas are no longer equal. The more different the inverters are the further away  $V_{inversion}$  shifts from the bisecting angle. Therefore, the locations of the intercept of the two transfer curves of the inverters are different for each cell.

If the SRAM starts up, it is in an undefined state and the inputs of the inverters are interpreted a logical zeroes. Due to that, both inverters try to load their output to logic one. Because there are differences at both transfer curves and both inverters are coupled, they influence each other by loading their outputs. As soon as one inverter reaches point  $V_{inversion}$  at its output, it defines the state of the SRAM cell. To put it in a nutshell, the inverter with the most pronounced switching performance defines the state after a startup. The value of this state is the so-called startup-value and is used to build the PUF. **Figure 1** illustrates this startup behavior. The upper green and lower turquoise curve corresponds to the inverter outputs, while both horizontal curves (red and blue) are point  $V_{inversion}$ , which is not on the bisecting angle.

The different startup-values are illustrated in **Figure 2**. The x-axis represents the voltage of the inverted SRAM output (output voltage of the inverter with the black solid curve) and whereby the y-axis indicated the voltage of the non-inverted output (output voltage of the inverter with the black dashed curve). Thereby three types can be distinguished:

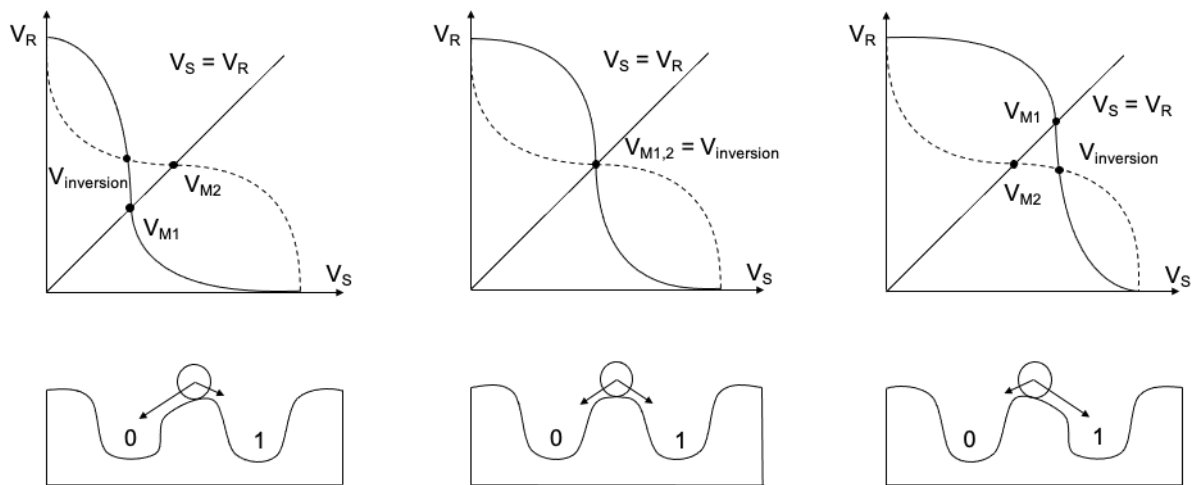
- $V_{inversion}$  is shifted to the left-hand side and the probability that the SRAM cell adopts Startup-Value zero is highest
- $V_{inversion}$  is shifted to the right-hand side and the probability that the SRAM cell adopts Startup-Value zero is highest
- $V_{inversion}$  is located on the bisecting and the probability for both Startup-Values are equal



**Figure 1:** The SRAM startup behavior by non-equal inverters

While manufacturing the SRAM, the attempt is made to manufacture both inverters identically. Due to that, the differences of both transfer curves are very slight and point  $V_{inversion}$  is very close to the bisecting angle. Furthermore, voltage fluctuations, described as Gaussian normal distribution with probabilities for bigger deviations, causes that cells with priorities for one particular startup-value adopt the non-preferred one. This behavior is the more important the closer  $V_{inversion}$  is to the bisecting angle. The startup-value adopted by the cells, can only be determined with a certain probability.

This startup-values behaviour is illustrated in **Figure 2**. The ball, in the lower part of **Figure 2**, symbolizes the current state of the SRAM cell. The inclined plane corresponds to the preference of  $V_{inversion}$ , whereby the two valleys represent the startup-values. By applying supply voltage, the ball is putted on the plane and rolls by its probability and preference, in the valley related to logical one or logical zero. This behaviour corresponds to the top left and top right illustrations. In the center of **Figure 2**,  $V_{inversion}$  is very close to the bisecting angle and do not have a strong preference. This results, that the ball rolls into valley logical one with the same probability than in logical zero.



**Figure 2:** Transfer curves of the SRAM Cells and their corresponding behaviour

Besides that, it is important to take time relevant aspects like capacitances into account. By those properties it is possible that the inverter with the weaker switching performance reaches  $V_{inversion}$  faster at its output, because its capacitances are not as big. Therefore, this weaker inverter is able to dominates and determinates the state of the SRAM cell. Another aspect, the capacitances cancel out the shift from  $V_{inversion}$  and cause a behaviour described by type three form the upper list.

### 3. Evaluating the SRAM behaviour: The examination setup

To evaluate the stability of the startup-values, it is necessary to examine a sufficient amount of SRAMs. In this work 30 SRAM are inspected and every startup-values is generated 500 times. This enables a statistic significantly result with a high resolution over the trend of the generated startup-values. Therefore, an evaluation platform is designed and implemented. The evaluation platform generates the startup-values of each SRAM, reads it and stores it at a memory card. Along each cycle, the platform measures und stores the ambient temperature as well. To avoid distortions due to residual voltage, it is advisable to use buffer time between every generation cycle and enable enough time for discharging before each new read cycle. In this case the delay is set to 100 milliseconds, which is given by the datasheet of the SRAM (Integrated Silicon Solutions, 2012). Other delays caused by the electronic components and wires have to be considered too.

The measurement series was done in a conventional office environment and lasted 8 days, 15 hours and 37 minutes. Figure 3 gives an insight to the setup. The heating of the room warmed the ambient temperature during the day to 21°C, while this is inactive at night. As a result, the temperature fluctuated between day and night by several °C. Since there is a window on the south-east side of the room, solar radiation was present in the morning and forenoon. Those influences are kept by the temperature sensor of the evaluation platform.



**Figure 3:** The evaluation platform during the measurement series

#### 4. Evaluation metrics

To determine fluctuations of measured values of a sample to its arithmetic mean, standard deviation is a common metric. It is determined by the square root of variance  $s^2$ . In contrast to the variance, which specifies deviation as the square of the origin unit, it remains the same for the standard deviation. According to Schiefer & Schiefer (2018)

$$s = \sqrt{s^2} \quad (1)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n x_i - \bar{x}^2 \quad (2)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

Furthermore, the range of variation  $R$  determines the difference between the biggest and lowest measured value.

$$R = x_{max} - x_{min} \quad (4)$$

A common deviation for technical applications is the logarithmic normal distribution. According to (Mathai & Haubold, 2017), the following applies to the density function with  $x < 0$  is:

$$f_x = \frac{1}{\sigma x \sqrt{2\pi}} \cdot e^{-\frac{\ln x - \mu^2}{2\sigma^2}} \quad (5)$$

and for the distribution function accordingly:

$$F(x) = \int_0^x f t dt \quad (6)$$

It is examined whether and in which way the measured values show a dependence. To do so, the statistical analysis methods of regression is used. The method uses the Gaussian minimization principle to best fit a function to the measured values. This is done by determine parameters of a given function. In this case it is beside the upper density function the following one:

$$xy = a \cdot x^b + c \quad (7)$$

A detailed discretion of this procedure is given Mathai & Haubold (2017). To determine how "well" the recession curve describes the behavior of the measurement values, the coefficient of determination  $B$  is a

common metric. It is the square of the correlation coefficient  $r_{x,y}$  and specifies the percentage of deviation which is determined by the curve. The correlation coefficient  $r_{x,y}$  is calculated by:

$$r_{x,y} = \frac{\sum_{i=1}^n x_i - \bar{x}y_i - \bar{y}}{\sqrt{\sum_{i=1}^n x_i - \bar{x}^2 \cdot \sum_{i=1}^n y_i - \bar{y}^2}} \quad (8)$$

And therefore, the coefficient of determination B is:

$$B = r_{x,y}^2 \quad (9)$$

The correlation is a degree for a linear relation of two characteristics. Its unit is, like the one of the coefficient of determination B, dimensionless. The range of its value is -1 to +1, where -1 indicates a negative and +1 a positive relation. According to Schiefer & Schiefer (2018)  $|r_{x,y}|$  holds:

- 0 no linear correlation,
- 0 bis 0,5 weak linear correlation,
- 0,5 bis 0,8 mean linear correlation,
- 0,8 bis 1,0 strong linear correlation, and
- 1,0 perfect correlation.

If no linear correlation is given, it does not mean, that there is no relation at all. A nonlinear relation is still possible, which can only be detected by considering the diagrams.

## 5. Trend of stability

The results are differentiated into two parts: the trend of stability and the distribution of the stability. If the start-up values change, their trend is similar to a root function. Changes in ambient temperature influences the behavior of the SRAM as well and the root function similar trend is modified. The diagrams **Figure 4** and **Figure 5**, illustrates the trend of the two representative SRAMs: SRAM 10 and SRAM 20. SRAM 10 has the lowest and SRAM 20 the highest correlation with their trend of changed startup values and the ambient temperature. All 30 SRAM are combinations of those by their correlation coefficient. Due to that, SRAM 10 and SRAM 20 are served as a case study. On the x-axis is the run of the measurement and on the y-axis is the absolute number of cells that have changed during the measurements. The red dashed curve is the regression curve, while the blue crosses are the actual measured values. The right secondary y-axis is the temperature in °C. This includes the green dots with the error bars, which correspond to  $\pm 1^\circ\text{C}$  and thus the inaccuracy of the DHT22 from the data sheet (Sertronics, 2018).

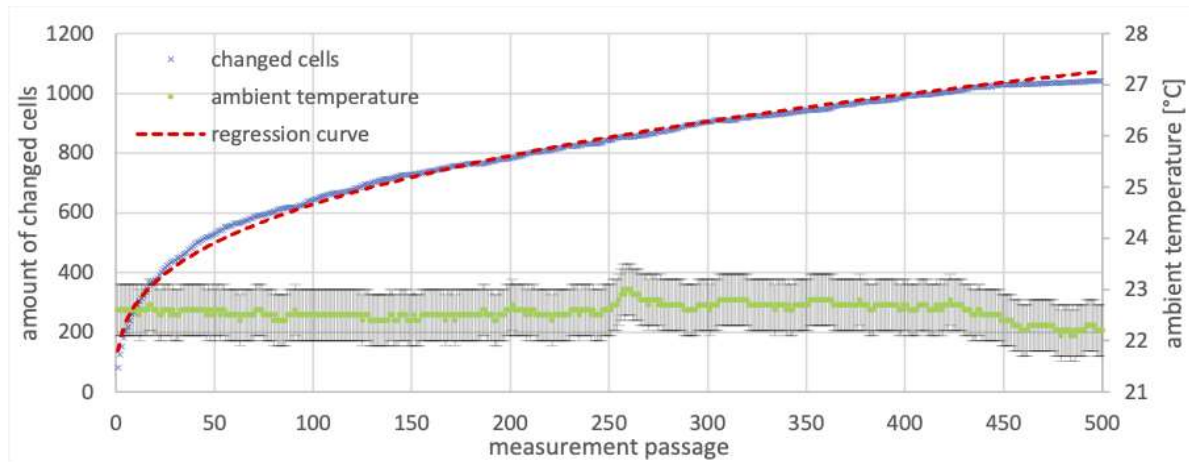
### 5.1 Trend of stability of SRAM 10

This SRAM has the lowest correlation with ambient temperature, resulting in  $r_{x,y}=0,0238$ . The regression curve is determinate to  $y(x) = 135,71 \cdot x^{0,3328}$  and has a coefficient of determination with  $B=0,9863$ . It can be seen that shortly after the temperature drop at the end of the course, a flattening of the measurement course takes place. The short-term temperature increase in the middle, on the other hand, has no influence. With a range of variation of  $R=0,9\pm 1^\circ\text{C}$ , the ambient temperature is relatively constant. Its average is  $(22,5644\pm 0,5)^\circ\text{C}$  with a standard deviation of  $s=0,1569^\circ\text{C}$ . The biggest difference with 1042 cells takes place at the end of the measurement. This corresponds to 0,0041% by which the SRAM is differently related to its first measurement. Despite the behavior, the very small maximum change of 0.004% is practically not relevant for the SRAM as a whole. The course can probably be traced back to the heating of the SRAM due to the readout. By increasing the temperature, electrical characteristics like electron mobility is changed. This results in a modification of the transfer curves of the inverters and therefor for point  $V_{inversion}$ , which has influences to the Startup-Values. (Jitty, et al., 2016) provide a study about the dependency of MOSFET and the temperature.

### 5.2 Trend of stability of SRAM 20

The highest correlation with the temperature given by  $r_{x,y}=0,9106$  has SRAM 20. The regressions function  $y(x)=0,1368x+8,9531$  is determinate with a coefficient of determination of  $B=0,9597$  and not as good as by SRAM 10. This dependency is recognizable in **Figure 5**. At the beginning, when the temperature is still constant, the same root curve is qualitatively similar to that of SRAM 10. Shortly after the temperature begins to rise steadily, the course of the measured values becomes much steeper. The ambient temperature rises from  $20\pm 0,5^\circ\text{C}$  at the beginning to  $22,4\pm 0,5^\circ\text{C}$  at the end of the measurement. The SRAM differs maximally by 82

Startup-Values and therefore by 0,0003% by the measurement 500. The ambient temperature has a linear relation with the trend of stability of the Startup-Values. In comparison to SRAM 10, this factor dominates and instead of the root function similar trend there is a linear one.

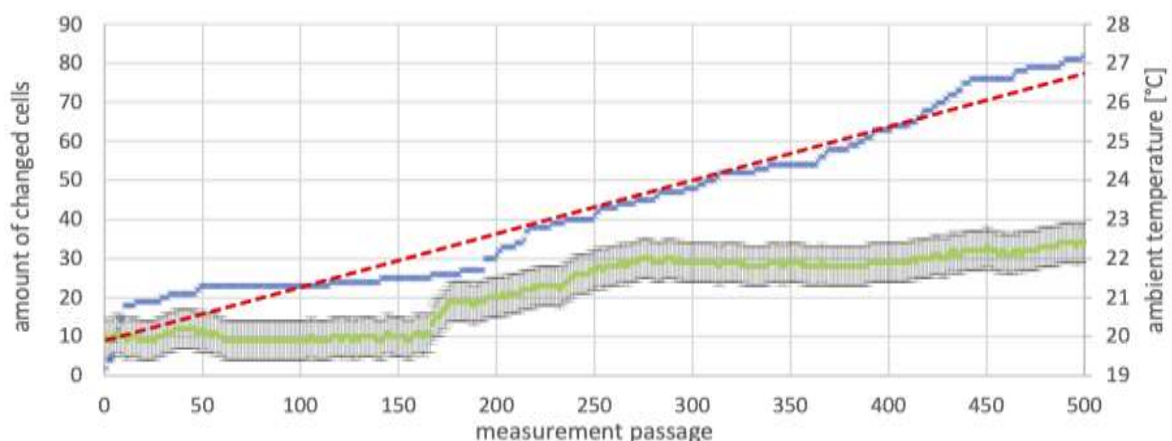


**Figure 4:** Stability trend of SRAM 10 with corresponding ambient temperature

This effect is caused by the heat flux from the SRAM to the ambient air. By changing the ambient temperature, the heat voltage increases or decrease whereby the heat dissipated more or less well. This results in a faster and higher increase in temperature in the SRAM. The property of thermodynamics is described by (Shankar, 2019). According to that, the behavior of Startup-Values is changed. The maximum change for the property of the SRAM practically irrelevant.

## 6. Distribution of the stability

The **Figures 6** and **8** illustrates, the stability of startup-values. The stability of the cells in percentage is given by the x-axis, while the percentage of cells that have at least reached stability is on the y-axis. Such a trend is called distribution function. This analysis is done, contrary to the previous one. According to this tool it is possible to use a self-constructed regression function. Due to that, a mirrored power function similar course can be determinate. The legend for the diagrams is the same as before. Since the stability has the same course mirrored on the y-axis due to its property in the range 0% to 50% and 50% to 100%, only the latter range is shown. The startup-values with a stability of 100%, are those whose point  $V_{inversion}$  is shifted far enough from the bisected angle away.



**Figure 5:** Stability trend of SRAM 20 with corresponding ambient temperature



### 6.1 Distribution of stability of SRAM 10

The trend given by **Figure 6** corresponds to a logarithmic normal distribution, whose density is shown in **Figure 7**. Since the range of 50% to 100% was chosen, the distribution and density is mirrored on the y-axis and shifted to the right by a factor of 1. The calculated parameters of the regression function are  $\sigma=10,87$  and  $\mu=46,17$ . Here, the coefficient of determination is  $B=0,9157$ . A stability of 100% have 99,58% of all cells. Since the data of the 500 measurements are used to determine the stability curve, this also contains the influence of the internal temperature increase by the read out process.

Comparing the number of cells that are not 100% stable (1054) with the number of cells that change maximally over the course of the measurement (1042), it is clear that there are 12 more unstable startup-values than those that change with the trend.

Those 12 cells are therefore cells whose point  $V_{inversion}$  is near enough to the bisected angle, so the startup-value is changing in a random way. (Lipps, et al., 2020) show that such a course can be expected even without the temperature influence. This is a typical distribution for technical applications and underlines the random production fluctuations in SRAM manufacture and the associated distribution of the Startup-Values. Thus, it cannot be assumed that the Startup-Values change only due to the internal temperature increase, but that the unstable cells from chapter Startup-Values contribute to this behavior.

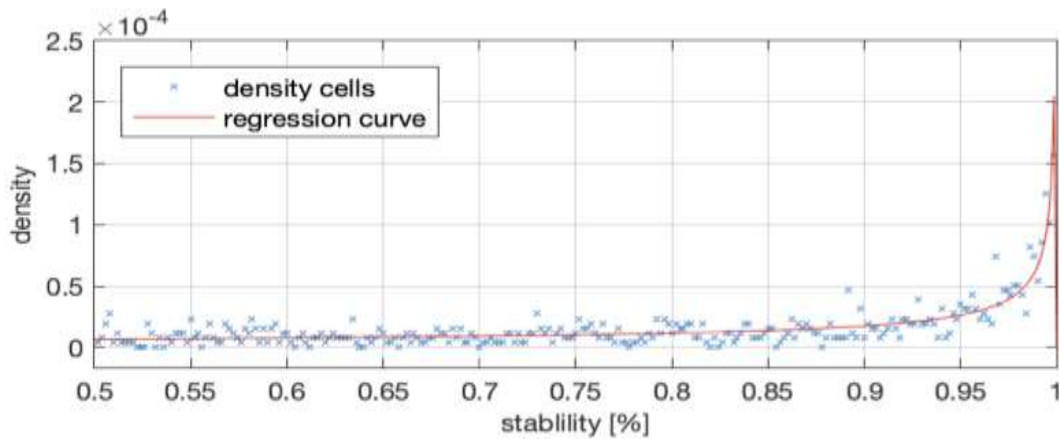


Figure 7: Density of stability of SRAM 10

### 6.2 Distribution of Stability of SRAM 20

There is a linear distribution of stability for SRAM 20 in **Figure 8**. The regression function is determined to  $y(x)=0,0005x+1$  by a coefficient of determination of  $B=0,987$ . 99,96% of all cells are 100% stable. The trend of the distribution of stability is dominated due to the high correlation of SRAM 20 and the ambient temperature ( $r_{x,y}=0.9106$ ) by it and has a linear behavior. According to this, there is no density distribution shown.

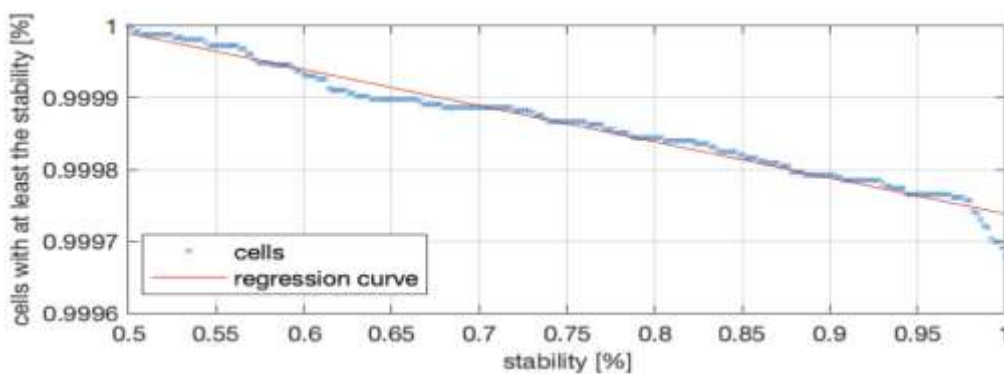


Figure 8: Distribution of stability of SRAM 20



## 7. Conclusion and future work

The work examines the appropriateness of SRAM-based PUFs for cryptographic applications in general. An evaluation is done by study of 30 SRAMs with the use of a self-developed platform. The results show that the stability of SRAM startup-values is very high. By considering the trends of stability, a root function similar curve is observable by a low correlation with the ambient temperature. The temperature influences the trend in that way that it is changed in a linear one. Overall, this qualitative behavior is recognizable, but in a global view the trend has not a practical influence. The same applies for the distribution of stability. The amount of 100% stable cells with over 99% is very high. The observable logarithmic normal distribution of SRAM 10 with the lowest correlation with the ambient temperature, is caused probably by internal temperature increase while the reading processes. Due to results of this work by testing 30 SRAM and each startup-value 500 times individually, it is proven that the stability to build a PUF is given.

The influence of ambient temperature requires further research. An approach is a repeat of the experiment in a temperature chamber by regulated ambient. Furthermore, the internal increase in temperature caused by the reading process and its influence is another important topic. This can be done by monitoring the internal temperature while testing the startup-values. There are also other possible ambient influences on the stability like humidity, air pressure, mechanical stress or radiation. In an industrial context for the IIoT, those are important and given conditions. How to build the PUF of a SRAM explicitly is a further challenge which needs research. A problem to do so is the deal with the non-stable cells even if they are few. By applying an error tolerance, so that those rare errors are tolerated, can be a solution. To avoid the decrease of security by implementing error tolerance, it is possible to add more startup-values to the PUF.

## Acknowledgements

This work has been supported by the Federal Ministry of Education and Research of the Federal Republic of Germany (Förderkennzeichen 16KIS0932, IUNO InSec and 01IS18062E, SCRATCh). The authors alone are responsible for the content of the paper.

## References

- Alheyasat, A., Torrens, G., Bota, S. & Alorda, B., 2020. Selection of SRAM Cells to improve Reliable PUF implementation using Cell Mismatch Metric. *XXXV Conference on Design of Circuits and Integrated Systems (DCIS)*.
- Böhm, C. & Hofer, M., 2013. *Physical Unclonable Functions in theory and Practice*. 1st Edition: Springer.
- Böhm, C., Hofer, M. & Pribyl, W., 2011. A Microcontroller SRAM-PUF. *5th International Conference on Network and System Security*.
- Bonneau, J., Herley, C., van Oorschot, P. C. & Stajano, F., 2012. *The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes*, IEEE Symposium on Security and Privacy.
- Chen, B., Ignatenko, T., Willems, F. M. J., Maes, R., van der Sluis, E. & Selimis, G., 2017. A Robust SRAM-PUF Key Generation Scheme Based on Polar Codes. *IEEE Global Communications Conference*.
- Farooq, M., Waseem, M., Khairi, A. & Mazhar, S., 2015. A Critical Analysis on the Security Concerns of Internet of Things (IoT). *International Journal of Computer Applications*, Volume 111, pp. 1-6.
- Gao, M. & Qu, G., 2014. *A highly flexible Ring Oscillator PUF.*, 51st ACM/EDAC/IEEE Design Automation Conference (DAC).
- Gao, Y., Su, Y., Yang, W., Chen, S., Nepal, S. & Ranasinghe, D. C., 2019. *Building Secure SRAM PUF Key Generators on Resource Constrained Devices.*, International Conference on Pervasive Computing and Communications Workshop.
- Gassend, B., Dwaine, C., van Dijk, M. & Srinivas, D., 2002. *Silicon Physical Random Functions.*, 9th ACM Conference on Computer and Communication Security, pp. 148-160.
- Guajardo, J., Kumar, S. S., Schrijen, G.-J. & Tuyls, P., 2007. *FPGA Intrinsic PUFs and Their Use for IP Protection.*, International Workshop on Cryptographic Hardware and Embedded Systems.
- Halak, B., 2018. *Physically Unclonable Functions - From Basic Design Principle to Advanced Hardware Security Applications*. 1st Edition, Southampton: Springer.
- Hoeller, A. & Toegl, R., 2018. Trusted Platform Modules in Cyber-Physical Systems: On the Interference Between Security and Dependability. *IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*.
- Integrated Silicon Solutions, 2012. *IS65C256AL 32K x 8 LOW POWER CMOS STATIC RAM.*: <https://cdn-reichert.de/documents/datenblatt/A300/IS62C256AL-ISSI.pdf>.
- Jitty, J., Ajith, R. & Keerthi K., N., 2016. Study of Temperature Dependency on MOSFET Parameter using MATLAB. *International Research Journal of Engineering and Technology*.
- Kusters, L., Ignatenko, T., Willems, F. M. J., Maes, R., van der Sluis, E. & Selimis, G., 2017. Security of Helper Data Schemes for SRAM-PUF in Multiple Enrollment Scenarios. *IEEE International Symposium on Information Theory (ISIT)*.
- Kusters, L., Rikos, A. & Willems, F. M. J., 2020. Modeling Temperature Behavior in the Helper Data for Secret-Key Binding with SRAM PUFs. *Conference on Communications and Network Security (CNS)*.

**Pascal Ahr, Christoph Lipps and Hans Dieter Schotten**

- Kusters, L. & Willems, F. M. J., 2019. *Secret-Key Capacity Regions for Multiple Enrollments With an SRAM-PUF*. Transactions on Information Forensics and Security.
- Lee, Y. S., Alasaarela, E. & Lee, H., 2014. Secure key management scheme based on ECC algorithm for patient's medical information in healthcare system. *The International Conference on Information Networking 2014 (ICOIN2014)*.
- Lipps, C., Ahr, P. & Schotten, H. D., 2020. How to Secure the Communication and Authentication in the IIoT: A SRAM-based Hybrid Cryptosystem. *European Conference on Cyber Warfare and Security (ECCWS)*.
- Lipps, C., Ahr, P. & Schotten, H. D., 2020. The PhySec Thing: About Trust and Security in Industrial IoT Systems. *Journal of Information Warfare*, 19(3), pp. 35-49.
- Lipps, C., Herbst, J. & Schotten, H. D., 2021. How to Dance Your Passwords: A Biometric MFA-scheme for Identification and Authentication of Individuals. *International Conference on Cyber Warfare and Security*.
- Lipps, C., Weinand, A., Krummacker, D., Fischer, C. & Schotten, H. D., 2018. *Proof of Concept for IoT Device Authentication based on SRAM PUFs using ATEMGA 2560-MCU*. South Padre Island, Texas, USA, 1st International Conference on Data Intelligence and Security (ISDIS-2018), pp. 36-42.
- Lipps, C., Mallikarjun, S. B., Strufe, M., Heinze, C., Grimm, C. & Schotten, H. D., 2020. Keep Private Networks Private: Secure Channel-PUFs, and Physical Layer Security by Linear Regression Enhanced Channel Profiles. *3rd International Conference on Data Intelligence and Security (ICDIS)*.
- Maes, R., 2013. *Physically Unclonable Functions - Constructions, Properties and Applications*. 1st Edition. Heverlee, Belgium: Springer.
- Mathai, A. M. & Haubold, H. J., 2017. *Probability and Statistics: A Course for Physicists and Engineers*. Berlin, Boston: De Gruyter.
- Ning, H., Farha, F., Ullah, A. & Mao, L., 2019. Physical Unclonable Function: Architectures, Applications and Challenges for Dependable Security. *The Institution of Engineering and Technology (IET)*, 14(4), pp. 407-424.
- Prada-Delgado, M. A. & Baturone, I., 2021. Behavioral and Physical Unclonable Functions (BPUFs): SRAM Example. *IEEE Access*, Volume 9, pp. 23751 - 23763.
- Rührmair, U., Hilgers, C., Urban, S., Weiershäuser, A., Dinter, E., Forster, B. & Jirauschek, C., 2013. *Optical PUFs Reloaded*.
- Schneier, B., 2015. *Applied Cryptography: Protocols, Algorithms and Source Code in C*. 20th Anniversary Edition ed. Indianapolis, IN, USA: John Wiley & Sons Ltd.
- Sertronics, 2018. *DHT22 - Digitaler Temperatur und Luftfeuchtesensor*.
- Shankar, R., 2019. *Fundamentals of Physics I: Mechanics, Relativity, and Thermodynamics*.: Yale University Press.
- van der Leest, V., van der Sluis, E., Schrijen, G.-J., Tuyls, P. & Handschuh, H., 2012. Efficient Implementation of True Random Number Generator Based on SRAM PUFs. In: D. Naccache, *Cryptography and Security: From Theory to Applications. Lecture Notes in Computer Science*: Springer, pp. 300-318.
- Wang, R., Selimis, G., Maes, R. & Goossens, S., 2020. Long-term Continuous Assessment of SRAM PUF and Source of Random Numbers. *Design, Automation & Test in Europe Conference & Exhibition (DATE)*

# Asylum Seekers From Russia to Finland: A Hybrid Operation by Chance?

Kari Alenius

University of Oulu, Finland

[kari.alenius@oulu.fi](mailto:kari.alenius@oulu.fi)

DOI: 10.34190/EWS.21.069

**Abstract:** This paper analyses interpretations made of the arrival of asylum seekers through Russia to Finland in the period from November 2015 to March 2016. Prior to that time, there were almost no asylum seekers at all from Russia to Finland, and their arrival ended just as abruptly in the spring of 2016. The news published in Finland's leading media during the above-mentioned period has been reviewed for the purposes of the study. Even as it was happening, different interpretations of the nature of the issue were presented in the media. The topic is important in three ways. First, the activities of the Finnish media in connection with this small-scale crisis have hardly been studied at all. It is therefore now possible to make a basic analysis of how Finnish media reacted to the surprising situation that had arisen on the border between Russia and Finland. Second, if it was a hybrid warfare operation, what Russia achieved through it should be evaluated. Third, the analysis can be used to assess the likelihood of a strong influx of refugees into the European Union via Russia. It is very well known that Russia, like other great powers, is interested in increasing its influence abroad. Russia is also ready to use a variety of means to achieve its goals. In particular, the conquest of Crimea and the outbreak of war in eastern Ukraine in 2014 have shown that Russia does not shy away from aggressive operations that combine traditional military means with modern unconventional practices where necessary. However, it cannot be concluded that whenever Russia seems to be carrying out operations planned by state leadership, they really are such. Each case deserves a separate analysis.

**Keywords:** hybrid operations, asylum seekers, refugees, Finland, Russia

---

## 1. Introduction

The main questions of the article are: What were the assumed reasons for the unexpected arrival of asylum seekers from Russia to Finland? What kind of reactions did it provoke in Finland? To what extent and on what grounds have conclusions been drawn as to whether this was a Russian operation? The strength of the conclusions presented in the media is assessed by comparing them with open research data available on Russia's activities in general.

The main source material of the study is the Finnish press. The movement was covered from 1 November 2015 to 31 March 2016, during which almost 1,800 refugees arrived in Finland. For the first ten months of 2015, and afterward, for the remainder of 2016, only a few asylum seekers arrived via Russia. The digitized collections of the National Library of Finland have been utilized in finding the material; "*Venäjä*" (Russia) and "*raja*" (border) have been used as search words. Of the later memoirs related to the theme, the work of Nina Järvenkylä (2019), in which she interviewed the Minister of the Interior Petteri Orpo and the Chief of Staff of the Ministry of the Interior Päivi Nerg, is worth mentioning. Orpo and Nerg were in the most important positions of responsibility at the Finnish Ministry of the Interior in 2015–2016.

The most important study published so far is the article by Arild Moe and Lars Rowe (2016). Their article was also based on media materials. The researchers' work focuses on events in Norway, but it also briefly discusses Finland. In autumn 2015, about 5,500 asylum seekers came from Russia to Norway. When Norway significantly tightened the conditions under which asylum seekers were admitted in November 2015, the flow of asylum seekers turned to Finland. Moe and Rowe's article and the interpretations presented in it are the most important point of comparison against which the events in Finland are examined in this paper.

It is clear that the authorities responsible for Finnish security have analysed the events of 2015–2016 in many ways. However, these reports and studies are not publicly available. Similarly, when the authorities responsible for security commented publicly on matters in 2015–2016, they did not tell everything they knew. Considering such a source-critical perspective is essential. There was no similar need to conceal things in the press, but on the other hand, there are other factors to consider in the media, such as the fact that the main goal of commercial media is to attract paying customers. Dramatic stories and interpretations sell. In addition, it must be noted that people working in the media have sympathies and antipathies, as do all people. Media channels also have political preferences.

## **2. “Driven by economic difficulties, exploited by the mafia and FSB”**

In November 2015, about 250 asylum seekers came to Finland from Russia. The number was very small compared to the total of about 30,000 refugees who had arrived in Finland in the summer and autumn. Most of them had come via northern Sweden (Finnish Immigration Service 2016). Against this background, it is understandable that refugees from Russia did not arouse much interest in the Finnish media during the first few weeks. The newspapers published only a few short news reports, which merely stated that a few refugees had also come from Russia (*Turun Sanomat* 4 Nov 2015; *Kaleva* 23 Nov 2015).

The situation and attitudes towards it began to change in early December. The number of asylum seekers from elsewhere began to fall sharply, but at the same time the number of refugees from Russia increased. At this point, concerns arose among the authorities and the media about how large an influx of asylum seekers from the East could grow in the future (*Uusi Suomi* 2 Nov 2015; *Yle Uutiset* 3 Dec 2015).

In early December, for the first time in the Finnish media, questions were asked about the reasons behind the changed situation. The most important public speech came from Major General Laitinen, Deputy Chief of the Finnish Border Guard. He commented that the reason was “financial difficulties”. In Finland, attention had already been drawn to the fact that a large proportion of asylum seekers were foreign workers who had lived in Russia for a long time, but now they had decided to seek asylum in Finland. Laitinen also pointed out that when Turkey started to block the entry of refugees into Europe, refugees needed to look for new routes (*Kaleva* 2 Dec 2015).

Moe and Rowe (2016) bring up the same explanatory factors and find them credible. Indeed, in autumn 2015, the European Union and Turkey had reached an agreement that Turkey would stop the influx of refugees into Greece (EU) via Turkey. This marked a sharp decline in the number of refugees from the Middle East. After that, some of the refugees tried to enter Western Europe by alternative measures. Crossing the Mediterranean from Libya was one way. Another, much longer but perhaps safer route was via Russia (Seitsonen et al 2017). In that case, Norway and Finland were the best-known and, due to their high standard of living, the most attractive destinations. Of Russia’s neighbours in the European Union, Estonia and Latvia were less interesting in this respect, and they received almost no asylum seekers via Russia (*Country Factsheet: Estonia 2016; Policy Report: Latvia 2016*).

It is also true that the situation of foreign workers residing in Russia had clearly deteriorated during 2015 as the Russian economy contracted. The EU and the United States had decided on economic sanctions against Russia, because of Russia’s onset of aggression against Ukraine during 2014. Therefore, many non-European migrant workers in Russia had lost their jobs and livelihoods (*Global Voices* 2016). In that situation, staying in Russia or returning home was not necessarily an attractive option. It probably seemed like a better option to try to get to Norway or Finland as an asylum seeker instead.

On this basis, Moe and Rowe (2016) suggest that the entry of asylum seekers into Norway and Finland was not a Russian hybrid operation but was a movement spontaneously initiated by the asylum seekers. Moe and Rowe admit that the transportation of asylum seekers to the Norwegian and Finnish border was organized, but that Russia’s state leadership was not involved. According to Moe and Rowe, it was the activities of the Russian mafia and corrupt lower-ranking officials, and it was in the interests of Russia’s state leadership rather to stop illegal operations.

The Finnish authorities and the media also pointed out that the mafia and local lower-level authorities arranged for the transport of asylum seekers. Representatives of the Finnish media interviewed the refugees and received detailed information about why they had gone to Finland, how much the trip had cost, what route they had taken, and with whom they had practically dealt (*Yle Uutiset* 8 Jan 2016; *MTV Uutiset* 23 Jan 2016). From this information, it is evident that the mafia and local officials did indeed play a key role in the practical organization. Likewise, it seems very likely that behind the phenomenon were the asylum seekers’ own motives. Subsequent research (Piipponen & Virkkunen 2020) has confirmed the information collected by the Finnish media. Moe and Rowe’s (2016) argument that this was not an operation carefully planned and launched by the Russian state leadership can thus be considered credible.

However, it seems that Moe and Rowe's interpretation is only valid when analysing where the movement began. Based on information published in the Finnish media, it can be concluded that the further the situation developed, the greater was the active role of Russia's highest leadership in the events.

### **3. "Finland at the mercy of Russia"**

At the beginning of January 2016, the assessment of the situation in Finland changed significantly. The first signal was President Sauli Niinistö's New Year's speech (Niinistö 2016). Niinistö said that "there may also be exploitation behind recent population movements that may be used as an instrument of power politics".

The anxious caution of high-level politicians and authorities characterized almost all their public speeches during the winter and spring of 2016. For example, Interior Minister Orpo and his close colleague Nerg categorically denied that Russia could have malicious motives (*Kauppalehti* 27 Jan 2016; *Demokraatti* 17 Mar 2016). The same caution and cover-up were repeated in the interviews published in 2019 (Järvenkylä 2019). However, in the Finnish media, the refugee question on the eastern border became one of the main topics of news and speculation around mid-January. For instance, *Kaleva*, the leading newspaper in northern Finland, published several editorials and expert articles (*Kaleva* 21 Jan, 23 Jan, 26 Jan, 30 Jan 2016). The same concern and the extensive coverage that followed was reflected throughout Finnish media (*Hufvudstadsbladet* 23 Jan 2016; *Demokraatti* 28 Jan 2016; *Yle Uutiset* 30 Jan 2016; *Suomen Kuvalehti* 19 Feb 2016).

When the active role of Russian state leadership was discussed in public, attention was drawn to several issues that referred to solutions other than those of refugees, the mafia, and local authorities. Representatives of the Finnish media were able to interview refugees and a few anonymous Russian authorities. (*Yle Uutiset* 8 Jan 2016; *MTV Uutiset* 23 Jan 2016). Based on these interviews, it was clear in that the FSB controlled the entry of asylum seekers very closely. The FSB decided who was allowed to travel to Finnish border stations and how many people received such permission each day. It is worth remembering here that the FSB was directly under Prime Minister Putin (Atlantic Council 2020).

Several specialists (Galeotti 2018; Dawisha 2014; Lucas 2012) in Russian politics have stated that the Russian state leadership works closely with organized crime, or one can speak of a kind of symbiosis. Russia's highest leadership controls the mafia, allows it to operate relatively freely, and can even order services from it. At the same time, the mafia undertakes to pay part of its profits to members of the political elite. This means that the arrival of asylum seekers in Finland in the autumn of 2015 was also an issue of which the FSB was aware and which it regulated at its discretion. The information presented in the Finnish media was thus fully in line with the existing research.

At the same time, the question was raised as to why asylum seekers had only just begun to come. This was not only about the increased number of refugees, but also about why Russia let them into Finnish border stations. Until 2015, there was an agreement between the Finnish and Russian authorities that Russian border guards would not allow persons to enter border stations without a visa to Finland. Moe and Rowe (2016) have noted that Russia had gradually reduced controls on those moving in border areas as early as the early 2010s. According to them, asylum seekers, the mafia and corrupt local officials simply took advantage of the situation in autumn 2015.

However, Moe and Rowe (2016) do not answer the question that bothered the Finns: why did asylum seekers come to only two border stations - Raja-Jooseppi and Salla - in northern Finland? Finland had a total of nine official border stations on the Russian border. Why did asylum seekers not come to the stations in central and southern Finland, which were much closer to St. Petersburg and Moscow? Thus, at seven border stations, Russian authorities maintained the practice of not allowing refugees to enter the Finnish border without a visa. At the two northernmost border stations, the practice had now changed unexpectedly, and refugees were transported to the border stations in a manner and with schedules regulated in detail by the FSB (*Ilta-lehti* 23 Jan 2016; *Apu* 3 Mar 2016).

The Finnish media described in detail the negotiations that took place between Finland and Russia from January to March 2016. Finland first tried to solve the problem with the Finnish Border Guard's commander trying to reach an agreement with his Russian counterpart, but to no avail. Next, Interior Minister Orpo met with the head of the FSB and the Russian Interior Minister, but no solution was reached. After Orpo, it was the turn of Prime

Minister Juha Sipilä, who met with Russian President Dmitry Medvedev, and again the negotiations ended without any progress (*Kauppalehti* 27 Jan 2016; *Helsingin Sanomat* 14 Feb 2016). Decisive talks took place between Finnish President Niinistö and Russia's real ruler, Putin, who was formally acting prime minister at the time, in mid-March (Järvenkylä 2019).

All Russian negotiators reiterated in public that the arrival of asylum seekers from Russia was spontaneous and that it was not organized in any way. The Russians also argued that according to international agreements, they did not have the right to prevent asylum seekers from seeking to enter Finland (*Kaleva* 30 Jan 2016; *Yle Uutiset* 31 Jan 2016) They did not comment at all on how it was possible that refugees arrived at only the two most remote border stations.

This peculiar phenomenon cannot be credibly explained except as a decision of the Russian state leadership. Russia's local border authorities did not have independent decision-making power in such a matter, but decisions at this level were always made by the Ministry of Internal Affairs (Giles 2019). The mafia would have had no reason to transport asylum seekers to only two border posts. Nor could asylum seekers have had any reason not to travel to the border posts that were closest. It is not known from open sources why Russia chose the two border posts mentioned above. At any case, no scholar, politician, or media representative has provided credible reasons to explain the phenomenon if it was a spontaneous action by asylum seekers or just a mafia decision without orders from the Russian state leadership.

There was a much more compelling reason for the caution of Finland's highest leadership than simply negotiation tactic. At the end of January, Brigadier General Kostimovaara of the Finnish Border Guard said in a rare blunt way that "Finland was at the mercy of Russia" (*Kaleva* 28 Jan 2016). The threat was that Russia would allow refugees to travel freely to Finland, or even start organizing it still more systematically. In that case, there could be hundreds of thousands of asylum seekers.

In February, the Russian state leadership showed that it could sovereignly influence the development of the situation. On February 26, Prime Minister Putin called on the FSB to "tighten controls on refugee flows from Russia to European countries". A couple of days later, the entry of asylum seekers into Finland ceased altogether, which means that the northern border stations also returned to the normal practice that had prevailed before (*Kaleva* 27 Feb 2016; *Turun Sanomat* 2 Mar 2016). The international agreements mentioned by President Medvedev and other senior authorities, which earlier "prevented Russia from interfering in the movement of refugees", were no longer an obstacle (Mikhailova 2018) and were not returned to in Russian public communication.

Putin did not refer to Finland in his order but spoke more broadly about "European countries" (*Yle Uutiset* 26 Feb 2016; *Kaleva* 27 Feb 2016). This was one fact that showed who Russia thought was the real opponent in the dispute. The same was pointed out by Medvedev, who criticized the "European Union's short-sighted immigration policy" (*Kaleva* 30 Jan 2016; *Yle Uutiset* 31 Jan 2016). According to Russia's message, the culprit was the EU. The real concerns were the economic sanctions imposed due to Russian aggression in Ukraine. In his public statements, Putin focused on this very issue, even though the negotiations with Finland formally concerned asylum seekers. Putin did not directly mention economic sanctions but spoke of economic relations at a more general level, though (*Yle Uutiset* 22 Mar 2016; *Kaleva* 23 Mar 2016).

The views expressed by the Finnish media are again supported by several researchers who have become acquainted with Russia's activities during the 2015–2016 refugee crisis. The interpretation of these scholars (Braghioli & Makarychev 2018; Nyquist & Cernea 2018) is that Russia sought to exploit or even escalate the refugee situation because it was in Russia's interests to undermine EU unity. The aim was to divert attention from the situation in Crimea and eastern Ukraine and to exacerbate disputes over the reception of asylum seekers in the EU. Several Mediterranean and Western European countries insisted that refugees should be distributed more evenly within the EU, but most Eastern European countries strongly opposed it. Striking a wedge between these parties increased the likelihood that the EU would not reach an agreement on the continuation of economic sanctions against Russia.

In addition, Giles (2019) and several other scholars (Jonsson 2019; Van Herpen 2014), for example, agree that Russia sees the European Union as its enemy in a deeper sense. It is not just about Crimea, Ukraine or economic sanctions, but about a seemingly permanent world view of the Russian state leadership. From the Kremlin's

perspective, the European Union (and the US) represents anti-Christianity and moral decay, while Russia is a defender of those values. The EU (and the US) is also seeking to plunder Russia's natural resources and subjugate Russia. Therefore, the Russian state leadership considers itself forced to take defensive measures and pre-emptive operations.

#### **4. Successes and failures of the Russian operation**

The above-mentioned statements from officials help outline what Russia was trying to do: the Finnish media wrote that Finland was a "prisoner of Russia" (*Kaleva* 30 Jan 2016; *Helsingin Sanomat* 4 Feb 2016). The exact content of the negotiations between Finland and Russia is secret. Still, it is almost certain that Russia tried to put pressure on Finland to work for the lifting, or at least easing, of economic sanctions. However, Finland succeeded in rejecting the demand that it act as Russia's agent within the EU. Even President Niinistö, who is cautious in his public statements, practically admitted what had happened. According to Niinistö, "Russia understands where Finland's limits go in the negotiations ... but everything outside the sanctions is free game in discussions" (*Kaleva* 24 Mar 2016).

Russia also stopped the organized import of asylum seekers to Finnish border stations. Did Russia fail in its operation, then? It is highly unlikely that Russia would give up completely in a situation where it had all the trump cards in its hand. The high-level Finnish representatives involved in the negotiations denied that the result would have been detrimental to Finland (*Kaleva* 23 Mar 2016; *Helsingin Sanomat* 24 Mar 2016). When the interviewer in 2019 asked Orpo what Russia got, Orpo's response was "nothing at all" (Järvenkylä 2019).

However, according to leaked information, the Finnish media were able to draw conclusions about what Russia got and how Finland had to bend. A memorandum from the Finnish Foreign Ministry revealed that Finland proposed restricting the use of the Raja-Jooseppi and Salla border stations so that they would be open only to citizens of Finland, Russia, the European Union, the EEA countries and Switzerland. This would have stopped the entry of asylum seekers altogether. According to Finland's proposal, the practice should also have been permanent (*Ilta-lehti* 12 Apr 2016).

The note sent by Russia to Finland and Finland's response to it show that Finland had to make two significant concessions. First, Russia agreed to conclude the agreement for only half a year. Russia would then decide whether it wanted to extend the agreement. Finland was thus still in "on probation", and correspondingly, Russia was free to think about how it could continue to put pressure on Finland. Another significant concession from Finland was that Russia excluded the EU and other Western countries from the agreement. Only citizens of Finland, Russia (and Belarus, which had a passport union with Russia) had the right to cross the border in Raja-Jooseppi and Salla. Finland had previously refused to make any bilateral arrangements with Russia that would have restricted the rights of other EU countries, so this was a legally significant case. However, Russia made this a condition of the entire agreement – only an agreement between the two states came into question (*Ilta-lehti* 12 Apr 2016).

Russia thus struck a political-legal wedge between Finland and other EU countries. At the same time, Russia was able to imply that it had succeeded in dragging its sphere of interest further west: Finland was equated with Belarus in this agreement. Russia was also able to set a precedent that it could refer to in the future. When Finland agreed to limit the competence of the EU and the rights of citizens of other EU countries in this matter, perhaps similar crisis situations would arise later, in which Finland would have to bypass the principles of normal conditions and conclude bilateral agreements with Russia. Based on the information available to the public, it is not possible to say whether Finland had to make any other secret concessions.

Can this episode in its entirety be called a Russian hybrid operation? The answer depends on how one wants to define "hybrid". There is no generally accepted definition of the term, but if the diversity of the means used is one of the essential criteria, then the answer may be in the affirmative. Of the leading Finnish politicians, only Minister of Defence Jussi Niinistö dared to say publicly that it was a Russian operation against Finland, and he specifically used the term "hybrid operation" (*Ilta-Sanomat* 29 Feb 2016).

Of the researchers of hybrid warfare, for example, Glenn (2009) has defined the matter as follows: "An adversary that simultaneously and adaptively employs some combination of (1) political, military, economic, social and information means, and (2) conventional, irregular, catastrophic, terrorism, and disruptive/criminal warfare

methods. It may include a combination of state and non-state actors". Veebel (2020), for his part, has aptly suggested that it is often better to talk about hybrid aggression rather than hybrid warfare.

It is also essential to note that the interpretations of Russia's readiness and willingness for various hybrid operations are based not only on counterparty interpretations. The matter can be read directly from Russia's military doctrine, which it has itself published. The matter was first brought to light by General Valery Gerasimov, Chief of the General Staff of the Russian Federation in 2013. According to him, modern warfare concentrates on the combined use of diplomatic, economic, political, and other non-military methods with direct military force, instead of waging an open war (Rác 2015).

## 5. Conclusions

The arrival of asylum seekers from Russia to Finland aroused great interest in the Finnish media. At the end of 2015, the media mainly discussed push and pull factors related to refugees in general. In January 2016, attitudes changed, and the issue began to be widely reported in the media as pressure on Finland, for which Russia's highest leadership was responsible. The perception created by the Finnish media was quite consistent.

All in all, what several well-known scholars have presented about Russia's geopolitical thinking and hybrid strategy is in line with what was written in the Finnish media about the refugee crisis of 2015–2016 on the Finland-Russia border. Of course, the media could only theorize based on open sources, public statements by the Russian state leadership, and interviews with refugees and anonymous Russian authorities. There is no demonstrable "smoking gun", but there are clear indications that the phenomenon of spontaneous refugee movement soon turned into a hybrid operation controlled by the Russian state leadership, in the practical implementation of which the Russian mafia played a significant role from start to finish. While this cannot be proven for sure, the most likely option seems to be that Russia decided to take advantage of asylum seekers who were in any case seeking the so-called "Arctic route" from the Middle East via Russia to Norway and Finland. Russia tried to put pressure on Finland to make political concessions, and it seems to have succeeded in part, although the breaking of the EU's sanctions front through Finland failed.

Through the operation, Russia was able to test the response of Finnish authorities, politicians and the media to this particular type of crisis, the development of which Russia was able to regulate. In addition, Russia was able to force Finland to conclude a border agreement in which Finland restricted the rights of other EU countries and their citizens. At the same time, forcing Finland into a bilateral agreement served as a principled message that Finland acknowledged its place on the list of countries where Russia could influence political decisions.

Through the operation, Russia also sought to underline the important role it played in curbing refugee flows to Europe. Russia showed that it can, if it so wishes, admit refugees to Europe, depending on whether Russia's wishes are taken into account in international arenas. Russia also announced through this small refugee crisis that it wanted to be treated as a great power with legitimate interests. It is obvious that Russia will continue to have similar opportunities for pressure. If Russia should decide to allow the free passage of asylum seekers to Europe, Finland, which has a 1,300-kilometre long common border with Russia, will not be able to prevent asylum seekers from entering Finland and the Schengen area.

## References

- Apu, 3 Mar 2016.
- Atlantic Council (2020), *Lubyanka federation: How the FSB determines the politics and economics of Russia*, <https://www.atlanticcouncil.org/in-depth-research-reports/report/lubyanka-federation/>.
- Braghirolli, S. and Makarychev, A. (2018), "Redefining Europe: Russia and the 2015 Refugee Crisis", *Geopolitics*, Vol. 23, Issue 4.
- Country Factsheet: Estonia (2016), [https://ec.europa.eu/home-affairs/sites/homeaffairs/files/08a\\_estonia\\_country\\_factsheet\\_2016\\_en.pdf](https://ec.europa.eu/home-affairs/sites/homeaffairs/files/08a_estonia_country_factsheet_2016_en.pdf).
- Dawisha, K. (2014), *Putin's Kleptocracy: Who Owns Russia?* Simon & Schuster, New York.
- Demokraatti, 28 Jan 2016.
- Demokraatti, 17 Mar 2016.
- Finnish Immigration Service (2016), [https://migri.fi/-/vuonna-2015-myonnettiin-hieman-yli-20-000-oleskelulupaa-uusia-suomen-kansalaisia-reilut-8-000#:~:text=Vuonna%202015%20Suomeen%20saapui%20turvapaikanhakijoita,%20ja%20syryriasta%20\(877\)](https://migri.fi/-/vuonna-2015-myonnettiin-hieman-yli-20-000-oleskelulupaa-uusia-suomen-kansalaisia-reilut-8-000#:~:text=Vuonna%202015%20Suomeen%20saapui%20turvapaikanhakijoita,%20ja%20syryriasta%20(877)).
- Global Voices (2016), "Russian Crisis Continues to Bite for Labour Migrants", <https://iwpr.net/global-voices/russian-crisis-continues-bite-labour-migrants>.



## Kari Alenius

- Galeotti, M. (2018), *The Vory: Russia's Super Mafia*, Yale University Press, New Haven and London.
- Giles, K. (2019), *Moscow Rules: What Drives Russia to Confront the West*, Brookings Institution, Washington.
- Glenn, R. (2009), "Thoughts on Hybrid Conflict", *Small Wars Journal*, <https://smallwarsjournal.com/blog/journal/docs-temp/188-glenn.pdf>
- Helsingin Sanomat*, 4 Feb 2016.
- Helsingin Sanomat*, 14 Feb 2016.
- Helsingin Sanomat*, 24 Mar 2016.
- Hufvudstadsbladet*, 23 Jan 2016.
- Ilkka-Pohjalainen*, 16 Feb 2016.
- Ilta-Sanomat*, 29 February 2016.
- Italehti*, 23 Jan 2016.
- Italehti*, 12 Apr 2016.
- Jonsson, O. (2019), *The Russian Understanding of War: Blurring the Lines Between War and Peace*, Georgetown University Press, Washington DC.
- Järvenkylä, N. (2019), *Tiukka paikka*, Docendo, Jyväskylä.
- Kaleva*, 23 Nov 2015.
- Kaleva*, 2 Dec 2015.
- Kaleva*, 21 Jan 2016.
- Kaleva*, 23 Jan 2016.
- Kaleva*, 26 Jan 2016.
- Kaleva*, 28 Jan 2016.
- Kaleva*, 27 Feb 2016.
- Kaleva*, 23 Mar 2016.
- Kaleva*, 24 Mar 2016.
- Kauppalehti*, 27 Jan 2016.
- Lucas, E. (2012), *Deception: Spies, Lies and how Russia Dupes the West*. Bloomsbury, London.
- Mikhailova, E. (2018), "Are Refugees welcome to the Arctic? Perceptions of Arctic Migrants at the Russian-Norwegian Borderland", Besier, G. and Stoklosa, K. (eds), *How to Deal with Refugees? Europe as a Continent of Dreams*, LIT Verlag, Wien.
- Moe, A. and Rowe, L. (2016) "Asylstrømmen fra Russland til Norge i 2015: Bevisst russisk politikk?", *Nordisk Østforum* 30 (2), <https://tidsskriftet-nof.no/index.php/noros/article/view/432/912>
- MTV Uutiset*, 23 Jan 2016.
- Niinistö (2016), *Tasavallan presidentti Sauli Niinistön uudenvuodenpuhe*, <https://www.presidentti.fi/puheet/tasavallan-presidentti-sauli-niiniston-uudenvuodenpuhe-1-1-2016/>.
- Nyquist, J.R. and Cernea, A. (2018), "Russian Strategy and Europe's Refugee Crisis", *Center for Security Policy*, [https://www.centerforsecuritypolicy.org/wp-content/uploads/2018/05/Russia\\_Refugee\\_05-28-18.pdf](https://www.centerforsecuritypolicy.org/wp-content/uploads/2018/05/Russia_Refugee_05-28-18.pdf)
- Piipponen, M. and Virkkunen, J. (2020), "The Remigration of Afghan Immigrants from Russia", *Nationalities Papers*, Vol. 48, Issue 4.
- Policy Report: Latvia (2016), *Policy report on migration and asylum in Latvia, reference year 2016*, [http://www.emn.lv/wp-content/uploads/APR\\_2016\\_part2\\_LATVIA\\_EN.pdf](http://www.emn.lv/wp-content/uploads/APR_2016_part2_LATVIA_EN.pdf).
- Rácz, A. (2015), *Russia's Hybrid War in Ukraine: Breaking the Enemy's Ability to Resist*, FIIA, Helsinki.
- Seitsonen, O., Herva, V. and Kunnari, M. (2017), "Abandoned Refugee Vehicles 'In the Middle of Nowhere': Reflections on the Global Refugee Crisis from the Northern Margins of Europe", *Journal of Contemporary Archaeology*, 3(2):244.
- Suomen Kuvalehti*, 19 Feb 2016.
- Turun Sanomat*, 4 Nov 2015.
- Turun Sanomat*, 2 Mar 2016.
- Uusi Suomi*, 2 Nov 2015.
- Van Herpen, M. (2014), *Putin's Wars: The Rise of Russia's New Imperialism*, Rowman and Littlefield, Lanham.
- Veebel, V. (2020), "Is the European Migration Crisis Caused by Russian Hybrid Warfare?", *Journal of Politics and Law*, Vol. 13, No. 2.
- Yle Uutiset*, 3 Dec 2015.
- Yle Uutiset*, 8 Jan 2016.
- Yle Uutiset*, 30 Jan 2016.
- Yle Uutiset*, 31 Jan 2016.
- Yle Uutiset*, 26 Feb 2016.
- Yle Uutiset*, 22 Mar 2016.

# Antarctica and Cyber-Security: Useful Analogy or Exposing Limitations?

Shadi Alshdaifat<sup>1</sup>, Brett van Niekerk<sup>2</sup> and Trishana Ramluckan<sup>2,3</sup>

<sup>1</sup>University of Sharjah, UAE

<sup>2</sup>University of KwaZulu-Natal, South Africa

<sup>3</sup>Educor Holdings, South Africa

[salshdaift@sharjah.ac.ae](mailto:salshdaift@sharjah.ac.ae)

[vanniekerkb@ukzn.ac.za](mailto:vanniekerkb@ukzn.ac.za)

[ramluckant@ukzn.ac.za](mailto:ramluckant@ukzn.ac.za)

DOI: 10.34190/EWS.21.013

**Abstract:** Antarctica is the last discovered continent, and is designated as a protected area for scientific research and peace and military operations are banned. Its status is governed by the 1959 Antarctic Treaty and subsequent agreements. These treaties related to Antarctica are cited as an example or a possible model for a treaty or international law for cyberspace. However, research facilities in Antarctica have fallen victim to cyber-attacks, and due to the environmental conditions, cyberattacks could potentially be devastating for the researchers who are affected. These cyber-incidents raise questions of how the current proposed application of international law for cyberspace will apply to such an area. The paper will assess the applicability of the Antarctica treaties to cybersecurity from two perspectives. The first perspective will discuss the existing proposals of adopting the treaties as a model for cyberspace, and the second will consider hypothetical scenarios based on previous cyber-security incidents in order to assess if the current proposals of international law are sufficient. These two perspectives will be contrasted to investigate the viability of the Antarctica Treaty system as a model for international treaties on cyberspace.

**Keywords:** area protection, cyber-attack, cyber-security, international humanitarian law, international security

---

## 1. Introduction

The US Department of Defense (2010: 37) states that “although it is a man-made domain, Cyberspace is now as relevant a domain for DoD activities as the naturally occurring domains of land, sea, air, and space.” This consideration of cyberspace as a domain for military operations has resulted in an increasing focus on the applicability of international humanitarian law (IHL) to cyber-operations. In 2019 both the Netherlands and France released national perspectives on how international law applies to cyberspace and cyber-operations (Ministère des Armées, 2019; Netherlands Parliament, 2019). Finland released their national perspective in 2020 (Ministry for Foreign Affairs, 2020). In addition, multi-stakeholder and non-governmental organisations have proposed norms for cyberspace, such as the *Paris Call* (2018) and Global Commission for the Stability of Cyberspace (GCSC, 2019).

Some authors have suggested the Antarctic Treaty System (ATS) as a model for laws and treaties for cyberspace; however, there are also challenges and concerns raised (Guarino & Iasiello, 2017; Maruhn, 2013; Stadnik, 2017; Ziolkowski, 2013a). In addition, Antarctica has suffered cyber-attacks, which will make this a unique case to discuss the applicability and limitations of international laws for cyberspace from the perspective of cyberspace as a global commons. This paper aims to discuss the suitability of the Antarctic Treaty System as a model for cyberspace by contrasting potential challenges that might arise from cyber-attacks against facilities in Antarctica.

The introduction continues in Section 1.1 with a background to Antarctica. Section 2 presents a background to the Antarctic Treaty System and a discussion on the proposal that a similar concept is used for cyberspace. Cyber-attacks related to Antarctica and their implications for international law are discussed in Section 3. A discussion on the two perspectives is presented in Section 4, followed by the conclusion in Section 5.

### 1.1 Background to Antarctica

The Antarctic Treaty Zone is defined as any territory south of 60°S Latitude (The Antarctic Treaty, 1959). The population within the zone can reach up to approximately 5000 researchers at research stations or vessels during the summer months; most of these are considered Internet users despite limited established telecommunications systems on the continent (CIA, 2020; Roberts, 2015). Antarctica is one of the harshest environments on the planet, and it is described as “the coldest, windiest and driest continent on Earth”

(Australian Antarctic Program, 2019). The coastal areas have an average temperature of -10 °C, ranging from 10 °C in summer to -40 °C in winter; the inland is colder with an average temperature of -60 °C, ranging from -30 °C in summer to -80 °C in winter (Australian Antarctic Program, 2019; Roberts, 2015). Winds can be sustained at 100km/h for days, and gusts can reach up to 200km/h (Australian Antarctic Program, 2019). Given these conditions, a cyber-attack against a research station or vessel could be potentially catastrophic should critical systems fail.

The communication infrastructure to the Antarctic research stations and vessels is limited, based primarily on satellite connections with the potential for limited private mobile phone connections near facilities. In 2015, the primary relay supporting 50Mbit/s was 90% of the total data capacity to the continent (CIA, 2020; Roberts, 2015). Despite this, Antarctica has an Internet domain: .AQ (Roberts, 2015).

## **2. Antarctica and International Law: Applications to Cyberspace**

### **2.1 The Antarctic Treaty System**

The Antarctic Treaty System (ATS) is the combination of the Antarctic Treaty (1959) and other agreements stemming from this. The Antarctic Treaty is interpreted as giving the continent the status of territory open for use in an unobstructed situation by any of the states, including those not among the parties to this treaty. This status permits the treatment of Antarctica as international territory, the legal status of which is similar to that of the high seas, air, or space. However, this is the main difference between the legal system of the Arctic and Antarctica. In so far, the treaty enshrined the right of states to exercise personal and territorial jurisdiction over potential territorial claims.

Since it is an international land and considered as a Common Heritage of Mankind, Antarctica should be treated as a peaceful land without the deployment of troop units, it cannot serve as a place for military operations, and to ensure environmental safety. However, Espach and Samaranyake (2020) indicate there is increasing tensions within the ATS, with some newcomers not following processes, and concerns that disputes over contested territorial claims may intensify.

Governance of the international commons is not customary; it is laid out in treaties including the 1967 Outer Space Treaty (NASA, 2006), the 1982 United Nations Convention on the Law of the Sea (UNCLOS), and to a lesser extent in the 1959 Antarctica Treaty System (Kriwoken & Keage, 1989). The governing treaty of Antarctica shares many similarities with UNCLOS as a protected international commons.

### **2.2 Using the Antarctic Treaty System as a basis for International Law for Cyberspace**

There are suggestions that Cyberspace should be considered as an area of global commons, similar to Antarctica (Guarino & Iasiello, 2017; Marauhn, 2013; Stadnik, 2017; Ziolkowski, 2013a). Currently, there are three such international spaces: Antarctica, outer space, and the high seas. For jurisdictional analysis, it is suggested that cyberspace should be treated as a fourth international space (Menthe, 1998). Cyberspace has no borders and does not fall under any one nation's sovereignty (Ayers, 2016).

However, as these areas are not subject to state sovereignty (Guarino & Iasiello, 2017; Marauhn, 2013), there needs to be some international cooperation and treaty to govern this, for which there is not yet any for Cyberspace (Marauhn, 2013). Shackelford (2017) discusses the concept of a global commons like Antarctica be used as a basis, and that all software or code that can be used for cyber-attacks be banned; however, he indicates that this will not be feasible as it will stifle innovation and the preparedness against cyber-attacks. Ziolkowski (2013a:167) suggests that nations should "precautionary measures concerning to potential cyber threats posing a significant risk of damage of a transboundary nature", similar to the *Protocol on Environmental Protection to the Antarctic Treaty of 1991*.

Inspections to verify compliance to *The Antarctic Treaty of 1959* are suggested in Article VIII(3) of the treaty; Kish (1995) suggested this implied espionage was permitted in Antarctica; however Ziolkowski (2013b) contests that from the various treaties the permissibility of espionage cannot be determined. Such inspections can be considered a mechanism to verify compliance with treaties in Cyberspace. However, the ease with which cyber-operations can be hidden raises concerns over the practicality of inspections related to Cyberspace; Shackelford (2017) notes the complexity for enforcement in Cyberspace.

Furthermore, the Antarctica treaty can be used as a basis for treaties for cyberspace since it is a field that always evolving. Therefore, it is crucial to keep in mind the fact that the lack of clear definitions concerning activities of cyberspace in Antarctica is comprehensible. For one thing, national appropriation appears to violate international law.

In short IHL provisions do not specifically mention cyber-operations. Because of this, and because the exploitation of cyber technology is relatively new and sometimes appears to introduce a complete qualitative change in the means and methods of warfare, it has occasionally been argued that IHL is ill-adapted to the cyber realm and cannot be applied in cyber-warfare (Dunlap, 2011). However, the absence in IHL of specific references to cyber-operations does not mean that such operations are not subject to the rules of IHL (Ferraro, 2015).

New technologies in the field of cyberspace are being developed, and international law is sufficiently broad to accommodate these developments. In specific, international law prohibits or limits the use of certain weapons specifically (for instance, chemical or biological weapons, or anti-personnel mines). But it also regulates, through its general rules, all means and methods of warfare, including the use of all weapons. In particular, Article 36 of Protocol I Additional to the Geneva Conventions (1979: 21) provides that: “[i]n the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party”.

It is conceivable to think that there is the potential for managing cyberspace within the context of international law. However, the principles of the Antarctica treaty become even more vital to be applied in cyberspace. As evidenced in the international community today, the current international law rules can be extended to cyberspace for the moment, as the use of cyberspace becomes more prevalent, specific provisions must be adopted to account for these differences and ensure that protections are sufficiently extended. Since cyberspace became a Common Heritage of Mankind (Baslar, 1998), however, this issue remains controversial (Alshdaifat, 2018).

All in all, one of the many phases required in the regulation of cyberspace is to address a uniform international regulatory system for such field, and that could be done through developing certain agreements in order for the states to abide by what we suggest is called a “cyberspace governance”.

We must emphasize that there is a need for a broader discussion on ways in which all international entities can participate in deep cyberspace endeavours, and coming to terms on this may require an agreement similar to the one established with the International Space Station Intergovernmental Agreement of 1998. In contrast to the ATS, the International Space Station Intergovernmental Agreement allows for the sovereignty of nations to extend to space for the portions of the International Space Station that they contributed (European Space Agency, 2021).

### **3. Cyber security and Antarctica**

#### **3.1 The potential for cyber-attacks against Antarctica**

Despite having limited infrastructure, connectivity and residents, Antarctica has experienced cyber-attacks. In 2003, the data acquisition and backup servers for the National Science Foundation’s Degree Angular Scale Interferometer radio telescope were compromised by a hacker; this was followed months later when data was stolen by Romanian hackers from the Amundsen-Scott South Pole Station in an extortion scheme. There was some confusion based on reports of the Amundsen-Scott South Pole Station hack whether the life support systems were compromised and if the researcher’s lives were in danger (Poulsen, 2004). In 2015 it was reported that three research stations in the Antarctic were compromised and used as staging points for cyber-criminals to target US and European government agencies (Roberts, 2015). In 2012 security firm Kratos was announced as winning the contract for the US Antarctic Program’s cybersecurity (InfoSecurity Magazine, 2012), indicating that cyber-security is being recognised as important for the facilities based on the continent. In addition, the US Antarctic Program has a webpage for their information security program, listing all the relevant information including awareness training and policies (USAP, c. 2020); this further illustrates the recognition of cybersecurity related to Antarctica.

Given the limited connectivity and redundancy of the connections, communications to Antarctica can be considered to be susceptible to distributed-denial-of-service (DDoS) attacks. As the above examples illustrate, there may be attempts to steal research data, particularly as there has been an increase in state-backed cyber-espionage operations, such as those targeting vaccine development during the COVID-19 pandemic (Cimpanu, 2020).

The increasing prevalence of cyber-physical attacks is also of concern. In 2009, safety systems were disabled in three oil rigs (Kravets, 2009), and navigations systems of both oil rigs and ships have been affected by malware (van Niekerk, 2016). Should a similar incident impact a research ship for the Antarctic programs, it may affect resupply of the bases and/or delay the departure of those at the bases. The concern of life support systems was mentioned above, and this concern is ratified by the disablement of the oil rig's safety systems. In the hostile environment of Antarctica, any negative impact on life support systems could have disastrous consequences.

The US National Oceanic and Atmospheric Administration implemented incident response measures in 2014 after a state-sponsored cyber-attack, which ultimately affected satellites, including weather satellites (Flaherty, Samenow & Rein, 2014). In 2020, a "hack-a-sat" competition was held at the Defcon conference, where participants would attempt to hack a satellite and the associated ground control (Scoles, 2020). By targeting satellites, a cyber-attack could disrupt communications or important weather information to the bases.

From the incidents described above, research bases in Antarctica can be, and have been, affected by cyber-attacks. These were cyber-criminals; however, there is scope for nation-state cyber operations to affect the bases and the researchers who reside there directly or indirectly by targeting the research vessels and/or satellites used for communications, navigation and weather information.

### **3.2 An International Humanitarian Law perspective on Antarctic cyber-attacks**

To apply IHL to cyber-attacks, a number of considerations should be taken into account: if the cyber-attack breached the sovereignty of a nation, if the attack affected the national processes (such as elections), or if there was equivalent damage or impact to that of a physical armed attack (Schmitt, 2017). Often, these considerations are difficult to determine, and the issue of attribution of the attack to another nation remains (Schmitt, 2017).

As the ATS indicates that there are no sovereign claims to regions in Antarctica, then a cyber-attack against a facility in the treaty zone cannot be considered to breach a nation's sovereignty. This clause also potentially precludes claims of interference as a cyber-attack against such facilities does not directly affect crucial national processes. These considerations will further complicate assessing the IHL implications of a cyber-attack and the appropriate responses. Should a cyber-attack result in significant physical damage, high risk to life or a loss of life due to critical life-support systems being disrupted, could this then be considered an armed attack or an act of war? In such a case, attribution to a nation-state will still be required, and it will be difficult to provide rapid digital forensic investigation in such a remote and hostile environment where there are limited transportation options and limited skill-sets at the location. Both the French and Dutch national perspectives consider the potential for a cyber-attack to be considered an armed attack or breaching the use of force if the consequences of the cyber-attack are similar to that of a physical armed attack (Ministère des Armées, 2019; Netherlands Parliament, 2019). Based on these interpretations, a cyber-attack on an Antarctic facility that results in death or injury of its occupants could conceivably constitute an armed attack.

Schmitt (2017) indicates that military cyber-operations in certain areas could breach treaties, specifically citing the ATS that prohibits military operations. A challenge would be attributing the cyber-attack against a facility in the Antarctic region to a nation's military organisation; in addition, the aggressor nation would need to be a signatory of the treaty in order for them to breach the treaty. A military cyber-operation could affect an Antarctic facility, but that facility may not be an intended target i.e. a digital equivalent to 'collateral damage'. Such an incident would then need to be assessed as to whether there was a military operation in the Antarctic, or if the military operation occurred elsewhere. In the latter case, the treaty would not be broken. The question of placing a military cyber-operation is then raised: is the cyber-operation deemed to be at the originating point, at the location of the targeted systems, or every point that the network traffic associated with the cyber-operation traversed?

A challenge also becomes jurisdiction in the event of a cyber-crime, as occurred in the incidents described in Section 3.1. If no nation has sovereignty, then who has jurisdiction to investigate crimes and enforce justice? This may then need to fall into a specialised multinational cyber-peacekeeping force, as proposed by Dorn and Webb (2019). Likewise, monitoring for compliance in cyberspace becomes more problematic, as malicious code can be easily transported via flash drive or other small storage devices. Therefore, the concept of physical inspections, such as for chemical, biological, or nuclear weapons, or for ensuring no military equipment enters the Antarctic zone, is not likely to be effective in cyberspace.

#### **4. Discussion**

The ATS provides for a protected space for scientific discovery, which can form the basis for an idealistic state for cyberspace. This may aid in the current uncertainties and shortcomings in applying IHL to cyberspace. However, there will need to be measures implemented to jointly 'govern' cyberspace.

A limitation to a treaty for cyberspace is that those who are not signatories to the treaty may still conduct cyber-attacks, and some signatories may still take the risk of cyber-operations due to the difficulty in attribution. This implies banning all weapons and offensive uses may disadvantage an injured state who abides by a treaty, as it may not have adequate means to respond to a cyber-attack, as is alluded to by Shackelford (2017). As the International Space Station Intergovernmental Agreement allows for the extension of sovereignty, considering aspects of agreements and treaties related to all common spaces may be more suitable than basing a cyberspace treaty on a single legal framework.

Should cyberspace be treated as a common space, there may be additional complexity in determining if a breach of IHL has occurred, specifically with regards to a nation's sovereignty. Jurisdictional problems may also arise in policing such a common space where everyone has access. Antarctica and outer space both are physical and have limited opportunity to access the physical location, however cyberspace is manmade and open to large portions of the population. This open access provides an opportunity for crime and states using proxies are greater in cyberspace. Therefore, monitoring and enforcement for treaty compliance will also need to be performed but will be more complex.

Given the perspectives discussed in this paper, there is some merit in designating cyberspace as a common space; however, this may be idealistic and additional complexities may arise as a result. It is unlikely that the ATS will be a suitable basis on its own, but in conjunction with other international agreements and treaties for common spaces, an adequate treaty for cyberspace as a common space could emerge.

#### **5. Conclusion**

The ATS is suggested as a possible model for a treaty in cyberspace. Despite its limited communications infrastructure, Antarctic facilities have been subjected to cyber-attacks. This paper aimed to discuss the applicability of the ATS to cyberspace from two perspectives: the existing proposals, as well as an assessment of applying existing IHL to the cyber-attacks against Antarctica. While there is some merit in using the ATS as a basis for a treaty in cyberspace, some limitations are also exposed. However, other international agreements for common spaces may aid in addressing these limitations. Therefore, the ATS cannot be used as a basis for a cyberspace treaty on its own, but all agreements and treaties for common spaces should be considered. It should be noted that cyberspace is different as it is manmade and there is broader access, whereas the existing three common spaces are naturally occurring physical spaces with limited access.

#### **References**

- Alshdaifat, S.A. (2018) "Who Owns What in Outer Space? Dilemmas regarding the Common Heritage of Mankind", *Pécs Journal of International and European Law* (vol II), 21-43.
- Australian Antarctic Program. (2019) Antarctic Weather, 18 February, [online], accessed 28 December 2020, <https://www.antarctica.gov.au/about-antarctica/weather-and-climate/weather/>
- Ayers, C.E. (2016) *Rethinking Sovereignty in the Context of Cyberspace*, Carlisle Barracks, Pennsylvania: U.S. Army War College. Available at: <https://www.hsdl.org/?view&did=802916>
- Baslar, K., (1998) *The Concept of the Common Heritage of Mankind in International Law*, The Hague: Martinus Nijhoff Publishers.
- CIA. (2020) Antarctica, *The World Factbook*, [online], accessed 4 January 2021, <https://www.cia.gov/the-world-factbook/countries/antarctica/>

- Cimpanu, C., (2020) "US formally accuses China of hacking US entities working on COVID-19 research", *Zero Day*, 13 May, [online], accessed 18 May 2020, <https://www.zdnet.com/article/us-formally-accuses-china-of-hacking-us-entities-working-on-covid-19-research/>
- Dorn, A.W., and Webb, S., (2019) "Cyberpeacekeeping: New Ways to Prevent and Manage Cyberattacks", *International Journal of Cyber Warfare and Terrorism* 9(1), 19-30.
- Dunlap, C.J., (2011) "Perspectives for Cyber Strategists on Law for Cyberwar", *Strategic Studies Quarterly* 5(1), 81-99.
- Espach, R., and Samaranayake, N., (2020), "Antarctica is the New Arctic: Security and Strategy in the Southern Ocean", *CNA*, 17 March, [online], accessed 12 March 2021, <https://www.cna.org/news/InDepth/article?ID=40>
- European Space Agency. (2021) International Space Station legal framework, [online], [https://www.esa.int/Science\\_Exploration/Human\\_and\\_Robotic\\_Exploration/International\\_Space\\_Station/International\\_Space\\_Station\\_legal\\_framework](https://www.esa.int/Science_Exploration/Human_and_Robotic_Exploration/International_Space_Station/International_Space_Station_legal_framework).
- Ferraro, T., (2015) "The ICRC's legal position on the notion of armed conflict involving foreign intervention and on determining the IHL applicable to this type of conflict", *International Review of the Red Cross* 97(900), 1227-1252. Available at <https://international-review.icrc.org/articles/icrcs-legal-position-notion-armed-conflict-involving-foreign-intervention-and-determining>
- Global Commission on the Stability of Cyberspace. (2019) *Advancing Cyberstability, Final Report*, November, [online], accessed 7 January 2021, <https://cyberstability.org/wp-content/uploads/2020/02/GCSC-Advancing-Cyberstability.pdf>
- Guarino, A., and Iasiello, E., (2017) "Imposing and Evading Cyber Borders: The Sovereignty Dilemma", *Cyber, Intelligence and Security* 1(2), 1-20.
- InfoSecurity Magazine., (2012) Kratos gets \$16 million cybersecurity contract for US Antarctic Program, 29 May, [online], accessed 22 November 2020, <https://www.infosecurity-magazine.com/news/kratos-gets-16-million-cybersecurity-contract-for/>
- Kish, J., (1995) *International Law and Espionage*, The Hague: Martinus Nijhoff Publishers.
- Kravets, D., (2009) "Feds: hacker disabled offshore oil platforms' leak-detection system", *Wired*, 18 March, [online], accessed 28 December 2020, <http://www.wired.com/2009/03/feds-hacker-dis/>
- Kriwoken, L.K., and Keage, P.L., (1989) "Introduction: the Antarctic Treaty System", In: J. Handmer (ed.), *Antarctica: Policies and Policy Development*, Centre for Resource and Environmental Studies, Canberra: Australian National University, 1-6.
- Marauhn, T., (2013) "Customary Rules of International Environmental Law - Can they Provide Guidance for Developing a Peacetime Regime for Cyberspace?" in: Ziolkowski, K. (ed.), *Peacetime Regime for State Activities in Cyberspace*, NATO CCDCOE: Tallinn, Estonia, pp. 465-484.
- Menthe, D.C., (1998) "Jurisdiction in Cyberspace: A Theory of International Spaces", *Michigan Telecommunications and Technology Law Review* 4(1), 69-103. Available at: <http://repository.law.umich.edu/mttlr/vol4/iss1/3>
- Ministère des Armées, (2019) *International Law Applied to Operations in Cyberspace*, [online], accessed 18 January, <https://www.defense.gouv.fr/content/download/567648/9770527/file/international+law+applied+to+operations+in+cyberspace.pdf>
- Ministry for Foreign Affairs, (2020) Finland published its positions on public international law in cyberspace, Government of Finland, 15 October, [online], accessed 2 November 2020, <https://valtioneuvosto.fi/en/-/finland-published-its-positions-on-public-international-law-in-cyberspace>
- National Aeronautics and Space Administration (NASA). (2006) *The Outer Space Treaty of 1967*, [online], accessed 25 January 2021, <https://history.nasa.gov/1967treaty.html>
- Netherlands Parliament, (2019) Appendix: International law in cyberspace, 26 September, [online], accessed 8 January 2020, <https://www.government.nl/binaries/government/documents/parliamentary-documents/2019/09/26/letter-to-the-parliament-on-the-international-legal-order-in-cyberspace/International+Law+in+the+Cyberdomain+-+Netherlands.pdf>
- Paris Call for Trust and Security in Cyberspace* (2018), 12 November, [online], accessed 18 January 2019, [https://www.diplomatie.gouv.fr/IMG/pdf/paris\\_call\\_cyber\\_cle443433.pdf](https://www.diplomatie.gouv.fr/IMG/pdf/paris_call_cyber_cle443433.pdf)
- Poulsen, K., (2003) "South Pole 'cyberterrorist' hack wasn't the first", *The Register*, 19 August, [online], accessed 22 November 2020, [https://www.theregister.com/2004/08/19/south\\_pole\\_hack/](https://www.theregister.com/2004/08/19/south_pole_hack/)
- Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, (1977), United Nations Treaty Series 3, 8 June, [online], accessed 9 January 2021, <https://www.refworld.org/docid/3ae6b36b4.html>.
- Protocol on Environmental Protection to the Antarctic Treaty (1991). Secretariat of the Antarctic Treaty: Buenos Aires, Argentina. Available at: <https://www.ats.aq/e/antarctictreaty.html>
- Roberts, P., (2015) "Petulant Penguin Attacks Use Antarctica As Base", *Security Ledger*, 1 April, [online], accessed 22 November 2020, <https://securityledger.com/2015/04/petulant-penguin-attacks-use-antarctica-as-base/>
- Schmitt, M.N. (2017) *Tallinn Manual 2.0: On The International Law Applicable to Cyber Operations*, Cambridge: Cambridge University Press.
- Scoles, S., (2020) "The Feds Want These Teams to Hack a Satellite—From Home", *Wired*, 6 August, [online], accessed 28 December 2020, <https://www.wired.com/story/the-feds-want-these-teams-to-hack-a-satellite-from-home/>
- Shackelford, S.J., (2017) "The Law of Cyber Peace", *Chicago Journal of International Law* 18(1), 1-47. Available at: <http://chicagounbound.uchicago.edu/cjil/vol18/iss1/1>

**Shadi Alshdaifat, Brett van Niekerk and Trishana Ramluckan**

- Stadnik, I., (2017) "What is an International Cybersecurity Regime and how we can Achieve it?" Masaryk University Journal of Law and Technology 11(1), 129-154.
- The Antarctic Treaty (1959), 1 December, Conference on Antarctica: Washington D.C. Available at: <https://www.ats.aq/e/antarctic treaty.html>
- United Nations Convention on the Law of the Sea (1982). 10 December, United Nations. Available at: [https://www.un.org/Depts/los/convention\\_agreements/texts/unclos/unclos\\_e.pdf](https://www.un.org/Depts/los/convention_agreements/texts/unclos/unclos_e.pdf)
- US Antarctic Program (c. 2020), USAP Information Security Program, [online] accessed 12 March 2021, <https://www.usap.gov/technology/sctninfosec.cfm>
- US Department of Defense, (2010) The Quadrennial Defense Review, February, [online], accessed 9 January 2021, <https://archive.defense.gov/qdr/QDR%20as%20of%2029JAN10%201600.pdf>.
- van Niekerk, B., (2016) "Analysis of Cyber-Attacks against the Transportation Sector", in: Korstanje, M.E. (ed.), *Threat Mitigation and Detection of Cyber Warfare and Terrorism Activities*, IGI Global: Hershey, PA, pp. 68-91.
- Ziolkowski, K., (2013a) "General Principles of International Law as Applicable in Cyberspace", in: Ziolkowski, K. (ed.), *Peacetime Regime for State Activities in Cyberspace*, NATO CCDCOE: Tallinn, Estonia, pp. 135-188.
- Ziolkowski, K., (2013b) "Peacetime Cyber Espionage – New Tendencies in Public International Law", in: Ziolkowski, K. (ed.), *Peacetime Regime for State Activities in Cyberspace*, NATO CCDCOE: Tallinn, Estonia, pp. 425-464.



# Evasion of Port Scan Detection in Zeek and Snort and its Mitigation

Graham Barbour, André McDonald and Nenekazi Mkuzangwe

Information and Cybersecurity Centre, Defence and Security Cluster, Council for Scientific and Industrial Research, Pretoria, South Africa

[gbarbour@csir.co.za](mailto:gbarbour@csir.co.za)

[amcdonald@csir.co.za](mailto:amcdonald@csir.co.za)

[nmkuzangwe@csir.co.za](mailto:nmkuzangwe@csir.co.za)

DOI: 10.34190/EWS.21.033

**Abstract:** East-west cyberattacks typically scan for open TCP ports on local network hosts in order to identify vulnerable services for subsequent exploitation. Since TCP port scans do not appear in legitimate network traffic, widely used intrusion detection systems such as Zeek and Snort include an option to search for these scans in background traffic, with the objective of alerting a network operator to potential threats. These port scan detectors are designed to trigger when the running count of rejected TCP connection attempts in a specified time interval exceeds a predetermined threshold (the rejection approach), and in the case of the Snort *sfportscan* pre-processor, when observing a dramatic increase in TCP connections over a relatively short period of time (the connection approach). In this paper, we present a novel algorithm for generating fast port scans that remain undetected by Zeek, thereby revealing a flaw that east-west cyberattacks may exploit. The differentiating factor of the new port scan algorithm is its rapid transmission of a spoofed TCP connection request to the network switch immediately after scanning each port. Port scans were conducted on a physical test network with an enterprise grade switch, where a network security testing and assessment appliance was used to generate background traffic from different application profiles. Experimental data is presented which demonstrates that (i) the novel scans remain undetected by Zeek for a scan rate of up to 1 million ports per second, and that (ii) neither Zeek nor Snort can detect the novel scans if the scan rate is reduced to 0.86 ports per minute or fewer. A strategy that combines the connection approach with a modified rejection approach for detecting the newly proposed fast port scans is proposed. It is concluded that this combined strategy holds potential for more reliable detection of port scans than the individual approaches. We envisage that the new port scan algorithm, the proposed detection strategy and the experimental findings would empower network security practitioners and designers of intrusion detection systems to address the shortcomings of existing detectors and improve detection strategies in general, thereby leading to more reliable detection of east-west cyberattacks.

**Keywords:** east-west cyberattack, port scan, TCP scan, packet spoofing, reconnaissance, intrusion detection, port scan detection, anomaly detection, Zeek, Snort

---

## 1. Introduction

The past two decades have witnessed a dramatic and sustained rise in cybercrime, with some cybersecurity experts predicting an annual loss of \$6.1 trillion USD to the global economy by 2021 (Morgan, 2020). Several factors have contributed towards the growth of cybercrime. These include global trends such greater levels of connectivity, the growing complexity of information systems, remote working and increased reliance on automation, as well as the ease with which malicious software tools can be obtained (Bayard, 2019) (Buil-Gil et al., 2020). Against this backdrop, technology for preventing cyberattacks and safeguarding information networks are becoming increasingly relevant.

Increasingly sophisticated malware and high volumes of phishing attacks globally have placed the so-called east-west cyberattack in the limelight (Crandall, 2020). In contrast to an external attacker attempting to penetrate the boundary of a network, at the start of the east-west cyberattack it is assumed that the attacker has already compromised at least one node on the intranet, thereby gaining remote access to the machine. These attacks involve the attacker gaining illegitimate access to additional nodes within the network, culminating in unlawful access to or modification of sensitive information and possible disruption of network services. Regardless of the ultimate goal, these attacks follow a series of distinct steps, namely reconnaissance, lateral movement and privilege escalation, and the exfiltration, corruption and disruption step.

In this paper, we focus on the initial reconnaissance phase of the east-west cyberattack (Shaikh et al., 2008). At the very start of the attack, the attacker has little or no knowledge of the network layout or vulnerable network services that may be exploited. Whereas the attacker may utilise passive means of conducting reconnaissance, such as inspecting the Address Resolution Protocol (ARP) table maintained by the local operating system, the information gained is limited in scope to the compromised host and the traffic it receives. To build a more comprehensive picture of the local network, the attacker proceeds with active reconnaissance, which involves

the use of software such as the popular Network Mapper (Nmap) utility (Lyon, 2008) to actively probe the network. Two distinct activities are identified, namely *sweep scanning*, where the attacker discovers whether a specified Internet Protocol (IP) address is being used by any node, and *port scanning*, where the attacker learns whether a specified Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) port is open on a specified IP address.

Reconnaissance presents an attractive opportunity for early detection of east-west cyberattacks as it precedes all other attack phases. Detection of the reconnaissance phase grants a longer time window for attack mitigation. The work presented here addresses TCP port scans and their detection. Several articles on detecting these scans have appeared in the literature, with the threshold-based detection strategy proving to be popular (Monowar, Bhattacharyya and Kalita, 2011). In this strategy, a detection is declared if one or more relevant traffic features exceeds a threshold inside a time window. In general, these methods may be divided into two groups according to the traffic parameters considered. We distinguish between *rejection-based* and *connection-based* approaches. In rejection-based detection, the number of rejected TCP connection attempts is considered as a feature, whereas the number of new TCP connections is considered in the connection-based approach. Both approaches have been successfully used in popular intrusion detection systems (IDSs) such as Snort (Roesch, 1999) and Zeek (Amann, 2021), formerly known as Bro (Paxson, 1999). In particular, the scan.zeek script follows the rejection-based approach, whereas the Snort *sfportscan* pre-processor module follows either the rejection-based or connection-based approach, depending on the detector sensitivity level selected by the user.

In this paper, we present a novel algorithm for conducting fast port scans – i.e., where a collection of server ports are scanned over a short period of time – that remain undetected by Zeek. This algorithm exploits a deficiency in Zeek that east-west cyberattacks in general may take advantage of. The differentiating factor of the new port scan algorithm is its rapid transmission of a spoofed TCP segment to the network switch immediately after sending the initial segment of the three-way TCP handshake to the server, where the latter segment is aimed at eliciting a response from the server that reveals whether the destination port is open or closed. The spoofed segment resembles the second stage of a successful TCP handshake, which is the expected reply from the scanned server in the case of an open port. In the case where the spoofed segment arrives at the detector before the true segment from the scanned node (which indicates a rejected connection attempt, if the port is closed), Zeek falsely declares the connection attempt a success. To demonstrate the viability of the proposed algorithm, tests were performed using a physical test network with an enterprise grade switch and a monitoring server running the Zeek and Snort intrusion detection software. The tests confirm that the new port scan algorithm can achieve rates of up to 1 million scanned ports per second without Zeek detecting the scan, whereas the Nmap synchronization (SYN) scan fails to evade Zeek at rates of 10 scanned ports per second and higher. In contrast, it was found that the spoofed TCP segment does not prevent Snort from detecting the new scan. Additional testing was conducted to evaluate the new algorithm’s sensitivity with respect to delayed arrival of the spoofed TCP segment at the monitoring node. These tests were conducted against a backdrop of network traffic generated by a Keysight PerfectStorm ONE network security testing appliance (Keysight, 2021). In general, it was found that the new port scan has a safety margin of 10  $\mu$ s for ensuring negligible probability of being detected by Zeek.

The remainder of this paper is set out as follows. In section 2, we provide background on TCP port scans and their detection. The Zeek port scan detection script is described in section 3 and the new fast port scan algorithm is presented in section 4. The setup of the test network and simulations is provided in section 5, whereas the test results are discussed in section 6. A strategy for mitigating the new port scan is presented in section 7, and conclusions are drawn in section 8.

## **2. Background**

This section describes the TCP port scan mechanism and different scan variants, as well as strategies for detecting these scans.

### **2.1 TCP port scanning**

TCP port scans are based on the TCP three-way handshake (Fall and Stevens, 2012), as illustrated in Fig. 1. A TCP connection attempt involves the client addressing a TCP segment with the SYN flag enabled to a server port. Depending on whether the port is open or closed, the server replies with a segment in which the SYN and acknowledgement (ACK) flags are enabled, or the reset (RST) and ACK flags are enabled, respectively. A third

possibility is that connection attempt times out; here, one or more of the segments do not reach their destination, or the server is configured not to respond if the port is closed (this is referred to as *port filtering*). Now consider the scenario where workstation “Malice” is to scan a TCP port on server “Alice” (refer to Fig. 2). Any TCP port scan involves Malice addressing a TCP segment to the relevant port on Alice, with the goal of eliciting a response from Alice that reveals the status of that port. Utilities such as Nmap (Lyon, 2008) declare a port as open, closed or filtered, where the latter term is used if the port status cannot be determined due to no segment being received from Alice.

TCP port scan variants differ with respect to the set of TCP flags that are enabled in the initial segment transmitted by Malice. The *TCP SYN scan* and *TCP connect scan* variants (Lyon, 2008) mimic the three-way TCP handshake and have several advantages compared to other scan variants. In particular, variants such as the *TCP Null*, *FIN*, and *Xmas* scans (Lyon, 2008) and the *Maimon* scan (Maimon, 1996) rely on the targeted server’s operating system strictly adhering to TCP RFC 793, and are in some cases unable to fully discern the true status of the targeted port. The newly proposed port scan follows the steps of the TCP SYN scan; in what follows, we summarise this scan variant and the closely related TCP connect scan. The *TCP SYN scan (or half-connect scan)* mimics an ordinary TCP connection attempt where Malice directly transmits a SYN segment to Alice, and waits a set period of time for a response from Alice (Lyon, 2008). The destination port is then declared open or closed if a SYN-ACK or RST-ACK segment is received from Alice, respectively. If no response is received in the time period, the port is declared as filtered. In the event of an open port, the connection is not completed by transmitting the concluding ACK segment of the TCP handshake, thereby leaving no trace of the attempt in Alice’s connection logs. This scan variant is arguably the most popular and widely used TCP port scan, due to its compatibility with any compliant server TCP stack, the high scan rate that can be achieved, and its reliability in discerning the true status of the port. The *TCP connect scan* is similar to the TCP SYN scan, except that the full three-way handshake is completed by Malice sending the concluding ACK segment (Lyon, 2008). This scan uses the *connect* operating system call to initiate the connection instead of writing raw packets to the interface, as is the case with the TCP SYN scan. Since the detection algorithms considered in this paper make use of features that are identical for both the TCP SYN scan and the TCP connect scan, we refer to these scans collectively as a *TCP SYN scan*.

## 2.2 TCP SYN scan detection

To facilitate the discussion of TCP SYN scan detection, the network diagram of Fig. 2 is considered, where workstation Malice again attempts to scan a port on server Alice, but with a port scan detector running on the monitoring node. In all cases, it is assumed that this node is connected to a mirror port on the switch, thereby granting it access to all traffic traversing the switch. The interested reader is referred to (Monowar, Bhattacharyya and Kalita, 2011) for a review of strategies for detecting the TCP SYN scan. Due to its use in the popular Zeek and Snort IDSs, we consider the thresholding strategy in this paper. In general, detectors that follow this strategy perform detection within a window of time referred to as the *detection window*. Within this window, the detector monitors one or more numeric traffic features that can be used for reliably discriminating between a TCP SYN scan and ordinary TCP connection attempts. The detector processes and combines these features before comparing them against a detection threshold, which may be static or adaptive. Any threshold crossing is considered to be anomalous and leads to a detection being declared within the window.

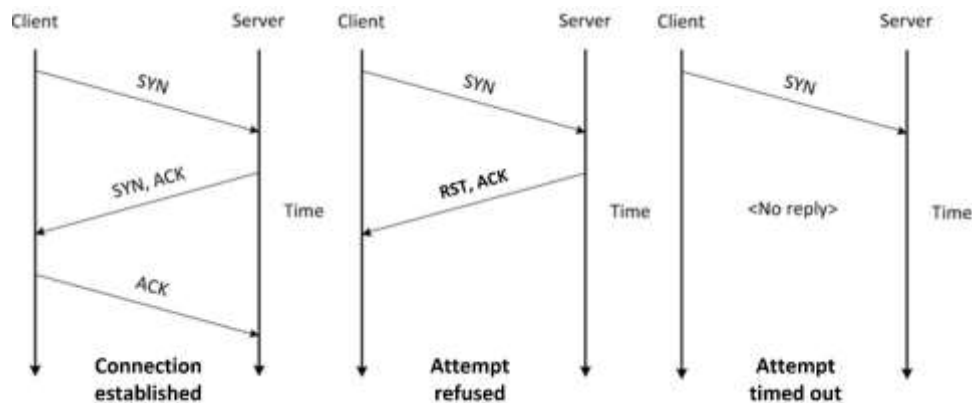
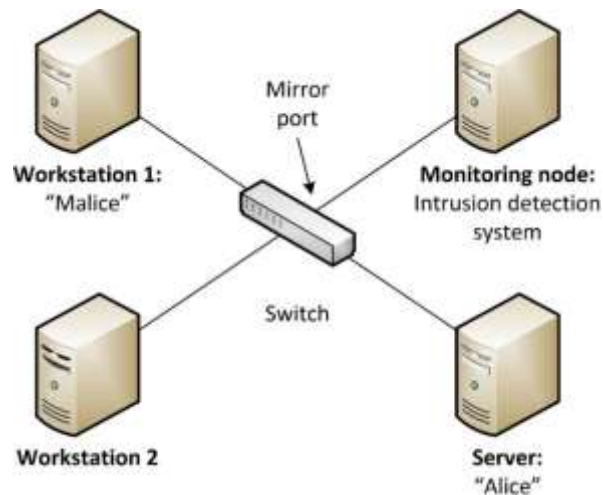


Figure 1: The TCP three-way handshake as illustrated with three scenarios, namely connection established (left), connection attempt refused (middle) and connection attempt timed out (right)



**Figure 2:** Network diagram for illustrating the TCP port scan mechanism

Detectors that follow the thresholding strategy may, in general, be divided into two groups according to the traffic feature that is monitored. We distinguish between the connection-based approach and the rejection-based approach.

### 2.2.1 The connection-based approach

The connection-based approach assumes that a sharp increase in the number of connection attempts from Malice corresponds to a TCP SYN scan. These detectors maintain a record of the number of TCP connection attempts initiated by each distinct machine, as characterised by the number of observed SYN segments within the detection window. A detection is declared upon this count exceeding the threshold.

The connection-based approach has the advantage of detecting TCP SYN scans that are sufficiently fast, regardless of whether the targeted ports are open or closed. However, this approach is subject to a trade-off between its detection rate (the fraction of port scans successfully detected) and false positive rate (the fraction of ordinary traffic misclassified as port scans), as brought about by the selection of the detection window length and the threshold. An important advantage of the connection-based approach is its ability to detect TCP SYN scans even if the targeted ports are filtered, in contrast to the rejection-based approach.

The Snort *sfportscan* pre-processor (Roesch, 1999) incorporates both the connection-based and rejection-based approaches to detection. Snort offers three detector sensitivity levels which determine the manner in which scan detection is conducted. When operating with a *low sensitivity level*, the detector tracks connection attempts from each distinct node over static 60 second detection windows, and maintains a count of rejected connections by searching for RST-ACK segment replies from Alice (the rejection-based approach). In contrast, when configured to operate with a *medium* or *high sensitivity level*, the detector follows the connection-based approach, and counts the number of connection attempts in the detection window. Experimentation with Snort (refer to section 6.1.2) has revealed that the high sensitivity setting mandates the use of a longer detection window than the medium sensitivity level, allowing Snort to capture slower port scans at the risk of a higher false positive rate. Whereas the fixed threshold level can be adjusted manually, it has been reported that scans destined to at least 5 distinct IP addresses or at least 20 different ports lead to a threshold crossing with the medium sensitivity level (Monowar, Bhattacharyya and Kalita, 2011).

### 2.2.2 The rejection-based approach

This approach is based on the assumption that Malice will necessarily scan many *closed ports* over a relatively short period of time. These detectors count the number of rejected TCP connection attempts originating from each node within the detection window. If any of these counts exceeds the threshold, a detection is declared and attributed to the corresponding node. The number of rejected connection attempts within the detection window is determined by *tracking* newly observed TCP connection attempts for a time period of at most  $T_C$  seconds, known as the *tracking window*. During this window, the detector waits for evidence that the destination port is open or closed. Upon receiving evidence, tracking is discontinued; alternatively, if no evidence is observed in the tracking window, the detector usually, but not always, considers the port closed. If the port is

deemed to be closed, the count of rejected TCP connections associated with the node that initiated the connection attempt is incremented. Rejection-based detectors differ with respect to what constitutes evidence of an open or closed port, as well as the specifics of counting rejections.

The TCP SYN scan (and SYN flood) detector of Korczynski et al. (2011) is a rejection-based system that considers a connection as accepted if the initial SYN segment from Malice is matched, within the tracking window, by *some* ACK reply from Alice. Otherwise, the connection is considered rejected. By requiring merely the existence of one such reply, the system requires only a sample of the connection data.

The Bro port scan detector (Paxson, 1999) considers a connection to be rejected if the connection attempt from Malice elicits an RST-ACK segment from Alice, or if no reply is received within the tracking period. The detector counts the number of distinct IP addresses against which rejected connections are directed. Jung et al. (2004) improved the Bro port scan detector by recording both accepted and rejected connections from each node within a window period, and computing a statistical detection metric based on sequential hypothesis testing.

We conclude this section by observing that the rejection-based detectors generally exhibit a significantly lower false positive rate than connection-based detectors. This is due to the fact that rejected connection attempts are relatively uncommon in typical network traffic containing no port scans.

### **3. The Zeek SYN scan detector**

Zeek provides the `scan.zeek` script for sweep scan and port scan detection. The following summary of the TCP SYN scan detection algorithm used in this script was compiled from the Zeek version 3.2.3 user manual (Amann, 2021) as well as experimentation by the authors.

The `scan.zeek` script follows the rejection-based detection approach, tracking new connection attempts over a tracking window in order to determine whether each port is open or closed. Experimentation revealed that

- 1. if the *first reply* from Alice within the tracking window is a SYN-ACK segment, Zeek immediately terminates further tracking of the connection, and the port is considered open;
- 2. if *the first reply* from Alice within the tracking period is an RST-ACK segment, connection tracking is immediately terminated and the port is considered closed; and
- 3. if the tracking period expires (neither of the previous two cases occurring), tracking is terminated and no action is taken, effectively treating such connections as legitimate.

A detection is declared if the number of unique ports that a particular node has failed to connect with exceeds a threshold within the detection window. Whereas both the detection window length and threshold can be configured by the user, Zeek specifies default values of 5 minutes and 15 ports for these parameters.

### **4. A novel algorithm for fast port scanning**

We present a novel algorithm for *fast* scanning of unfiltered ports while evading TCP SYN scan detection by Zeek. The algorithm exploits a deficiency of the `scan.zeek` script. This deficiency, as well as the manner in which it is exploited, is outlined using the same network configuration considered in section 2.2 (see Fig. 2).

Let Malice address a SYN segment to a closed port on Alice. Suppose that immediately after transmitting this SYN segment, Malice *spoofs* Alice's SYN-ACK reply (the segment that would be transmitted by Alice if the port were open) and directly transmits this segment to the switch. Note that Malice has access to the source and destination MAC and IP addresses of both the compromised node (workstation 1) and Alice, and can therefore immediately generate the header of the spoofed segment.

Now consider the TCP segments arriving at the monitoring node via the mirrored port. The initial SYN segment from Malice arrives first and Zeek enters the tracking window for this connection attempt. One of two cases then occurs. If the legitimate RST-ACK segment from Alice arrives first, case 2 of section 3 applies; the port is considered closed, and the count of rejected connection attempts for Malice is incremented. On the other hand, if Malice's spoofed SYN-ACK segment arrives first, case 1 of section 3 applies; the count of rejected connection attempts is *not* incremented, and the detection metric remains unaffected. If Malice transmits the spoofed segment *immediately* after transmitting the original SYN segment, it is reasonable to expect (and verified by simulation in section 6) that the spoofed segment has a high likelihood of arriving on the mirrored port *before*

the legitimate segment from Alice. In this case, Zeek ignores the latter segment, as it has already declared the connection legitimate after observing the spoofed segment. Thus Malice is able to evade Zeek. Note that the spoofed segment does not enter Alice’s TCP stack, and is therefore “invisible” to Alice. Consequently, Alice replies with the RST-ACK segment as it normally would, and Malice is in a position to view this response. Malice is therefore able to discern that the port is closed.

The description above assumed that the new port scan was directed against a closed port of Malice. The same sequence of events occur if the port were open, except that Alice now transmits a SYN-ACK segment, which Malice recognizes as proof of the targeted port being open. Therefore, the scan is successful when directed against open or closed ports.

### 5. Test network and simulation setup

The test network that was used to evaluate the scan algorithms is presented in Fig. 4. The network consists of a single broadcast domain with nodes connected to a 48 port Dell Force-10 S60 Gigabit Ethernet switch (FTOS version 8.3.3.8). Malice and the monitoring node are both physical servers running Ubuntu Linux 18.04.4 LTS, whereas the client and server pools consist of virtual machines hosted on a 1 Gbps Keysight PerfectStorm ONE appliance (Keysight, 2021). Client and server traffic were each multiplexed onto a single 1 Gbps Ethernet link. Keysight BreakingPoint software (version 8.10.1, ATI 285362) was used to generate network application traffic between the client and server pools. Three different application traffic profiles were used, namely a *corporate profile* (traffic from a typical business network), a *streaming profile* (traffic containing video streaming) and a *file sharing profile* (traffic with client-server and peer-to-peer file transfers). The traffic distribution for each profile is presented in Table 1, whereas the number of clients and servers is presented in Table 2.

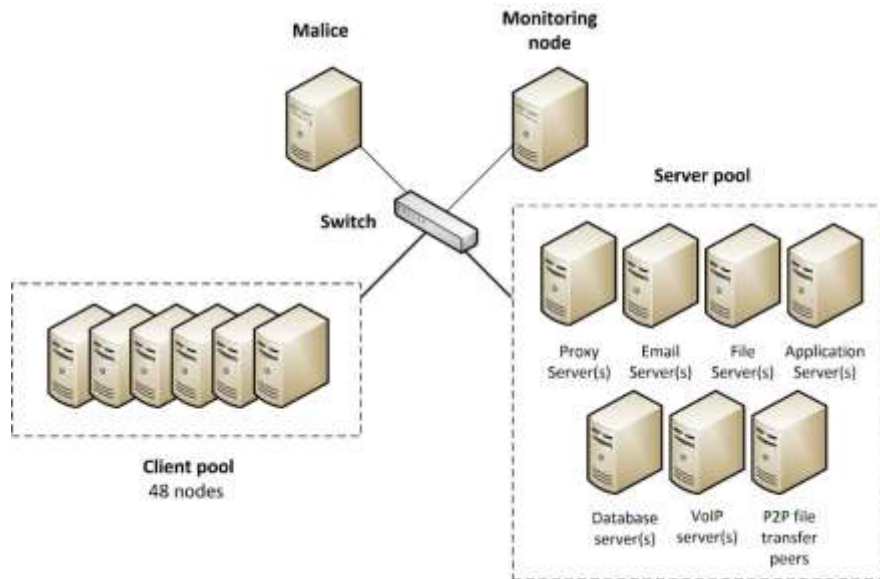


Figure 4: Diagram of the test network used for evaluating the port scan and detection algorithms

Table 1: Distribution of traffic for the corporate (C), streaming (S) and file sharing (F) profiles

Traffic type	Percentage of total bandwidth			Traffic subtype or protocol	Percentage of total bandwidth		
	C	S	F		C	S	F
HTTP/HTTPS	42%	37%	14%	Video	24%	31%	7%
				Audio	4%	4%	2%
				Text	14%	2%	5%
Email	14%	5%	5%	SMTP	14%	5%	5%
Client-server file transfers	12%	--	24%	FTP	6%	--	16%
				Dropbox	6%	--	8%
Peer-to-peer file transfers	--	--	40%	Bittorrent	--	--	40%
Database access	16%	--	--	MS SQL	8%	--	--
				PostgreSQL	8%	--	--
Office applications	6%	5%	7%	Office 365 Excel	1.5%	1.25%	1.75%

Traffic type	Percentage of total bandwidth			Traffic subtype or protocol	Percentage of total bandwidth		
	C	S	F		C	S	F
				Office 365 Onenote	1.5%	1.25%	1.75%
				Office 365 Powerpoint	1.5%	1.25%	1.75%
				Office 365 Word	1.5%	1.25%	1.75%
Voice over IP	10%	5%	10%	SIR/RTP	10%	5%	10%
Video Streaming	--	48%	--	Netflix	--	36%	--
				Hulu	--	12%	--

Tests involving port scan detection made use of Snort version 2.9.17 and Zeek version 3.2.3. The software was installed on the monitoring node, which was connected to a mirror port on the switch. Both the new algorithm and Nmap version 7.90 were used to conduct TCP SYN scans.

**Table 2:** Clients and servers used for generating corporate (C), streaming (S) and file sharing (F) profile traffic

Clients / server type	Count		
	C	S	F
Clients	48	48	48
Proxy servers	1	2	1
Mail servers	1	1	1
File servers	2	0	2
Application servers	1	1	2
Database servers	1	0	0
Voice over IP servers	1	1	1
Peer-to-peer clients	0	0	5
Total:	55	53	60

## 6. Results

Two sets of tests were performed to evaluate the novel port scan algorithm.

### 6.1 Test set 1: Evasion of Zeek and Snort port scan detection

The purpose of the first test set is to evaluate the novel scan algorithm's success in evading Zeek and Snort at various scan rates (i.e., ports scanned per second). Each test was repeated using Nmap instead of the new scan algorithm. All of these tests were carried out without generating any background traffic, thereby excluding false positives. Each scan was directed against unfiltered and closed ports on a server. The novel algorithm was configured to transmit the spoofed SYN-ACK segment immediately after transmission of the SYN segment.

#### 6.1.1 Port scan detection using Zeek

A total of 10 000 closed ports were scanned at rates of between 10 and 1 million ports per second. Table 3 reveals that the new algorithm is able to successfully evade detection up to the maximum scan rate of 1 million ports per second. In contrast, Nmap fails to evade the detector at each of the scan rates considered.

#### 6.1.2 Port scan detection using Snort

Tests involving Snort were conducted using a slower port scan rate of between 0.5 and 60 ports per minute, with a total of 32 ports scanned. Snort was configured to operate at low, medium and high sensitivity levels. Since Snort does not have the same deficiency as Zeek, evasion by spoofing is not possible. However, Table 4 reveals that a progressively slower scanning rate leads to successful evasion of Snort. In particular, Snort is unable to detect scans from either algorithm at rates below 3 ports per minute (low sensitivity), 6 ports per minute (medium sensitivity) and 0.86 ports minute (high sensitivity). The table also reveals that both scans are identical in the rate required to evade Snort, which implies that the spoofed TCP segment is not intercepted by Snort and used to aid detection.

### 6.2 Test set 2: Safety margin with respect to delay of spoofed segment arrival

The novel scan's success in evading Zeek depends on the arrival of the spoofed segment at the monitoring node before the RST-ACK segment from the scanned server. Factors that may delay the delivery of the spoofed

segment, and hence the segments' order of arrival, include the scanning node's hardware, the switching hardware and network traffic and server loads. The purpose of test set 2 is to investigate the novel scan's sensitivity with regards to these factors by simulating a network with realistic traffic profiles and conducting scans of closed ports. The tests from set 2 were conducted against a backdrop of traffic from each of the traffic profiles (Table 1), with the novel scan directed at closed ports on each of the servers in the test network (Table 2). During each scan, the arrival sequence of the segments was logged and processed to estimate the probability of Alice's RST-ACK segment arriving first at the monitoring node (this is a measure of *failing* to evade Zeek). A variable time delay  $T_\delta$  was introduced between Malice's transmission of the initial SYN segment and the spoofed SYN-ACK; by varying this parameter, the safety margin available to the algorithm was quantified.

**Table 3:** Results for detection by Zeek, where "S" and "F" denote a success or failure in evading the detector

Scan rate (ports per second)	10	10 <sup>2</sup>	10 <sup>3</sup>	10 <sup>4</sup>	10 <sup>5</sup>	10 <sup>6</sup>
Delay between scanning ports	100 ms	10 ms	1 ms	100 us	10 us	1 us
New port scan algorithm	S	S	S	S	S	S
Nmap SYN scan algorithm	F	F	F	F	F	F

**Table 4:** Results for detection by Snort, where "S" and "F" denote a success or failure in evading the detector

Scan rate (ports per minute)	60	20	12	6	4	3	2.4	2	1	0.86	0.75	0.67
Scan delay (seconds)	1	3	5	10	15	20	25	30	60	70	80	90
New scan, Snort low sensitivity	F	F	F	F	F	S	S	S	S	S	S	S
Nmap, Snort low sensitivity	F	F	F	F	F	S	S	S	S	S	S	S
New scan, Snort med. Sensitivity	F	F	F	S	S	S	S	S	S	S	S	S
Nmap, Snort med. Sensitivity	F	F	F	S	S	S	S	S	S	S	S	S
New scan, Snort high sensitivity	F	F	F	F	F	F	F	F	F	S	S	S
Nmap, Snort high sensitivity	F	F	F	F	F	F	F	F	F	S	S	S

The probability of the RST-ACK arriving before the spoofed SYN-ACK, as a function of the transmit delay  $T_\delta$  in seconds, is plotted in Fig. 5 to Fig. 7 for the corporate, streaming and file sharing traffic profiles, respectively. Each figure contains plots corresponding to mean sustained background traffic rates of 250 and 750 Mbps (left and right subfigures). A similar trend is observed over all traffic profiles, servers and traffic rates. The figures reveal that the probability of the RST-ACK segment arriving first decreases slowly to around 0.1 as  $T_\delta$  is decreased to 100  $\mu$ s (the *plateau* region). Further decreasing  $T_\delta$  below 30  $\mu$ s results in a rapid decrease (the *asymptotic* region), with a probability of  $10^{-4}$  reached at a transmit delay of 20  $\mu$ s. The steep slope suggests that a transmit delay no longer than 10  $\mu$ s would be sufficient to ensure a negligible probability of having the RST-ACK segment arriving first, thereby leading to successful evasion of Zeek.

## 7. Mitigation

Zeek may be modified to detect the novel scan. Consider the arrival of a new SYN segment from Malice at the monitoring node, such that Zeek enters the tracking window for that connection. Instead of the three cases considered in section 3, the modified version of Zeek now waits for the arrival of an RST-ACK segment from Alice within the tracking window. If this is observed, tracking is immediately terminated and the port is deemed closed. Alternatively, if no RST-ACK segment from Alice is observed by the end of the tracking window, no action is taken and the port is considered open. This modified algorithm detects all scans detected by Zeek, as well as our new scan. However, the modification incurs a computational burden of having to track legitimate connection attempts over the entire tracking window. We conclude this section by observing that the modified algorithm still cannot detect port scans directed against filtered ports. One may consider a strategy where the output of a connection-based detector is combined with the output of the modified detector of Zeek, such that a threshold crossing of either results in a detection. This combined approach can detect filtered port scans, but at the risk of increasing the number of false positives; this is due to nodes such as proxy servers and domain name servers often leading to bursts of connection attempts that are legitimate. Thus, careful selection of the connection-based detector's detection window length and threshold level is necessary, and the system would need to be configured to ignore active hosts. The combined strategy also has the potential to detect slower port scans directed at closed ports at a lower false positive rate than what is possible using a connection-based detector alone. Slower port scans can be detected by selecting a longer detection window for the rejection-based detector than the connection-based detector. Since rejected connections are relatively infrequent in network traffic free from scans, this would not result in the same increase in false positives as would be observed were a longer window selected for the connection-based detector.



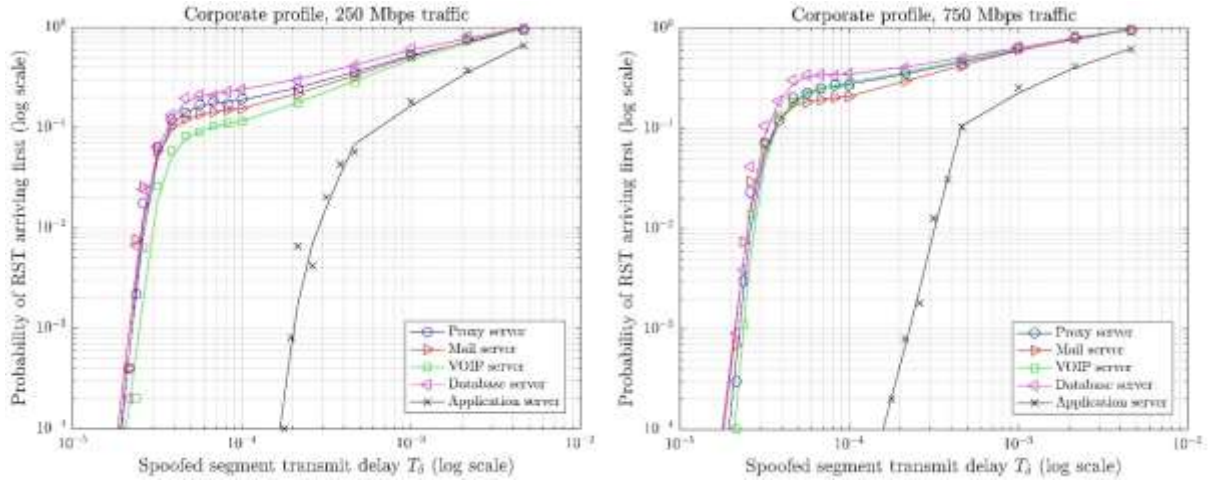


Figure 5: Probability of the RST-ACK segment arriving before the spoofed SYN-ACK segment at the monitoring node, with traffic generated from the corporate profile at a mean rate of 250 Mbps (left) and 750 Mbps (right)

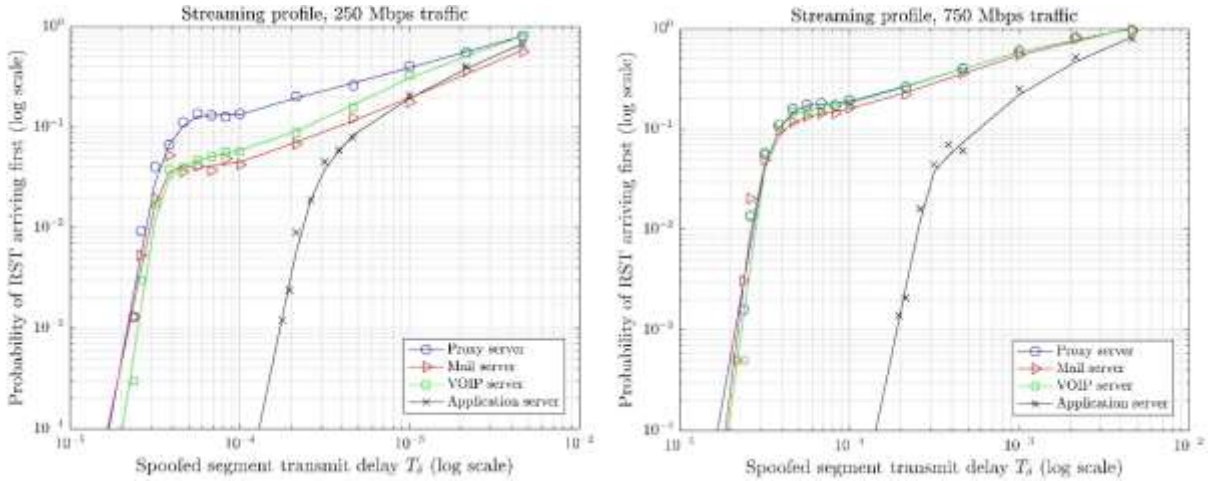


Figure 6: Probability of the RST-ACK segment arriving before the spoofed SYN-ACK segment at the monitoring node, with traffic generated from the streaming profile at a mean rate of 250 Mbps (left) and 750 Mbps (right)

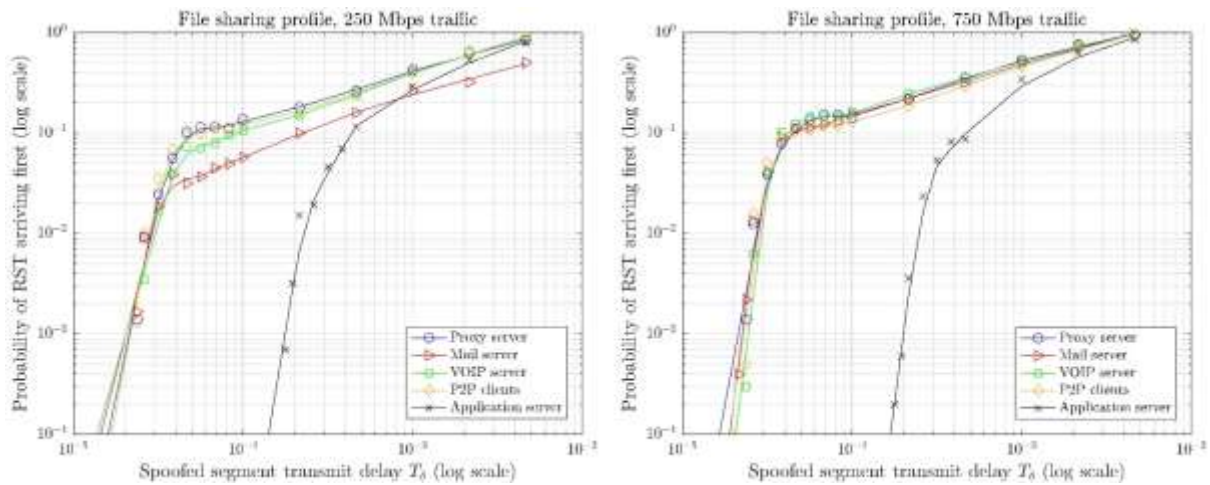


Figure 7: Probability of the RST-ACK segment arriving before the spoofed SYN-ACK segment at the monitoring node, with traffic generated from the file sharing profile at a mean rate of 250 Mbps (left) and 750 Mbps (right)

## 8. Conclusion

In this paper, a new algorithm for conducting fast TCP port scans was proposed. The novel scan uses TCP segment spoofing to exploit a deficiency in the Zeek port scan detection algorithm, thereby evading detection by this IDS. Our tests indicate that the new scan successfully evades Zeek while achieving rates of up to 1 million ports per second. The algorithm was also found to have a safety margin of 10  $\mu$ s delay with respect to factors that may affect the timely delivery of the spoofed segment at the IDS. We conclude that the proposed algorithm is a viable option for fast and stealthy reconnaissance in networks protected by the Zeek IDS, thereby emphasising the need to address the deficiency of the Zeek scan detector. A modification to Zeek that allows for detection of the new port scan was proposed. In addition, an approach whereby the modified detector is combined with a connection-based detector was considered. Whereas this combined detector has potential for more reliable port scan detection than the individual detectors alone, its performance needs to be verified through testing in operational networks.

## References

- Amann, J. (2021) "Zeek Documentation - Book of Zeek", [online], The Zeek Network Security Monitor, <https://docs.zeek.org/en/master/>
- Bayard, E. (2019) "The Rise of Cybercrime and the Need for State Cybersecurity Regulations", *Rutgers Computer & Tech. LJ*, Vol. 45, pp. 69-95.
- Buil-Gil, D., Miro-Llinares, F., Moneva, A., Kemp, S. and Diaz-Castano, N. (2020) "Cybercrime and Shifts in Opportunities During COVID-19: A Preliminary Analysis in the UK", *European Societies*, Vol. 23, pp. S47-S59.
- Crandall, C. (2020) "Want stronger cybersecurity? Start by improving east-west traffic detection", [online], GCN: The Technology that Drives Government IT, <https://gcn.com/articles/2020/09/24/east-west-traffic-monitoring.aspx>
- Fall, K. and Stevens, R. (2012) *TCP/IP Illustrated*, 2nd ed., Vol. 1, Addison-Wesley, Upper Saddle River, New Jersey.
- Jung, J., Paxson, V., Berger, A., and Balakrishnan, H. (2004) "Fast portscan detection using sequential hypothesis testing", Proceedings of the IEEE Symposium on Security and Privacy, pp. 211-225.
- Keysight Technologies, (2021) "Perfectstorm ONE | Keysight", [online], <https://www.keysight.com/zz/en/products/network-test/network-test-hardware/perfectstorm-one.html>
- Korczynski, M., Janowski, L. and Duda, A. (2011) "An Accurate Sampling Scheme for Detecting SYN Flooding Attacks and Portscans", Proceedings of the IEEE International Conference on Communications, pp. 1-5.
- Lyon, G. F. (2008) *Nmap network scanning: The official Nmap project guide to network discovery and security scanning*, 1st ed., Insecure Com LLC.
- Maimon, U. (1996) "TCP Port Stealth Scanning", *Phrack Magazine*, Vol. 7, No. 49.
- Monowar, B., Bhattacharyya, D. and Kalita, J. (2011) "Surveying Port Scans and Their Detection Methodologies", *The Computer Journal*, Vol. 54, No. 10, pp. 1565-81.
- Morgan, S. (2020) "Special Report: Cyberwarfare In The C-Suite", [online], Cybercrime Magazine, <https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021/>
- Paxson, V. (1999) "Bro: A System for Detecting Network Intruders in Real-Time", *Computer Networks*, Vol. 31, No. 23-24, pp. 2435-2463.
- Roesch, M. (1999) "Snort: Lightweight Intrusion Detection for Networks", Proceedings of the 13th Systems Administration Conference (LISA '99), Vol. 99, No. 1, pp. 229-238.
- Shaikh, S., Chivers, H., Nobles, P., Clark, J. and Chen, H. (2008) "Network reconnaissance", *Network Security*, Vol. 2008, No. 11, pp. 12-16.

# The Manifestation of Chinese Strategies Into Offensive Cyberspace Operations Targeting Sweden

Johnny Bengtsson<sup>1,2</sup> and Gazmend Huskaj<sup>3,4</sup>

<sup>1</sup>Swedish National Forensic Centre (NFC), Swedish Police Authority, Linköping, Sweden

<sup>2</sup>Department of Electrical Engineering, Linköping University, Sweden

<sup>3</sup>Department of Military Studies, Swedish Defence University, Stockholm, Sweden

<sup>4</sup>School of Informatics, University of Skövde, Sweden

[johnny.bengtsson@polisen.se](mailto:johnny.bengtsson@polisen.se)

[gazmend.huskaj@fhs.se](mailto:gazmend.huskaj@fhs.se)

DOI: 10.34190/EWS.21.066

**Abstract:** The aim of this article is to present how Chinese strategies are manifested into offensive cyberspace operations targeting Sweden. It is commonly known that People's Republic of China (PRC, and in this definition the meaning of the government and its military), uses five-year plans (FYP) for social and economic steering strategy of their country. This has been going on since 1953 until today. In 2015, the national strategic plan Made in China 2025 (中国制造2025) was launched by Le Keqiang, the Premier of the State Council of PRC. The main goal with this plan is to strengthen the economic development. In addition, Chinese military strategists noted the importance of information warfare and intelligence during military operations. This article is based on open sources: the official English translated version of the 13th Five-year plan (FYP) and other reporting on cyberspace operations linked to the PRC. A number of cases are presented to highlight the link between the PRC FYP and their targets. Next, the current situation in Sweden is presented and how the country is targeted by PRC-linked activities, both in and through cyberspace, but also military infiltration on academia. The results show that Sweden has been, and is continuously the target of offensive cyberspace operations. In parallel, the country is also the target of military infiltration on the academia, and direct investment strategies such as Huawei attempting to compete for the 5G frequency actions arranged by the Swedish Post and Telecom Authority. In conclusion, Sweden will continue to experience cyberespionage from PRC on all levels and on all domains; science, technology, IP and privacy information theft. Previously unveiled cyberspace operations cases in this article have proven to be a convenient strategy for the PRC to reduce its research and development gap in several ways; innovatively, financially and to shortening the time-to-market (TTM).

**Keywords:** Chinese strategies, cyberespionage, information warfare, offensive cyberspace operations, Sweden

---

## 1. Introduction

Chinese offensive cyberspace operations targeting western industries have increased and are aggressive. Collecting information through open sources, numerous Chinese cyberspace operations have been revealed and traced back to the People's Republic of China (PRC). These operations have all been conducted by the government, the military, "hacker" groups – or advanced persistent threat (APT) groups – supported or facilitated by the government or military of the PRC. This paper and its analysis will mainly focus on the state-sponsored APTs. The aim of this article is to present how Chinese strategies are manifested into offensive cyberspace operations targeting Sweden. The incentives for the unveiled governmental and military related cyber activities are directly or indirectly linked to PRC's underlying political, economic and military mid-term and long-term strategies regarding the 13th FYP, MIC2025 and other strategic documents for the wealth development of PRC and its citizens. Previous scholars agree that cyber operations may be used for espionage purposes to steal intellectual property and for economic espionage purposes (Harknett & Smeets, 2020; Thornton-Trump, 2019). In addition, "Chinese cyber-espionage operations are known to go back since at least the early 2000s" (Harknett & Smeets, 2020, p. 18).

The main contributions are summarized as follows:

- 1. Chinese strategies and their manifestation into offensive cyberspace operations are presented;
- 2. based on 1), their targets in various sectors are identified and presented;
- 3. finally, the conclusions present how the strategies are manifested targeting Sweden.

This study begins with a review of 2. Chinese Strategies, followed by 3. Chinese Information Warfare and the Evolvement of Cyber Doctrine and 4. Chinese Offensive Cyberspace Operations. In 5. The General Situation in Sweden is presented, followed by Conclusions in 6. Any acronyms used are presented in 8. Appendix 1.

## **2. Chinese strategies**

This section presents the Chinese strategies. It begins with the 13<sup>th</sup> Five-year plan (13<sup>th</sup> FYP), followed by Made in China 2025.

### **2.1 The 13th Five-year plan (13th FYP)**

It is commonly known that People's Republic of China (PRC, and in this definition the meaning of the government and its military), uses five-year plans (FYP) for social and economic steering strategy of their country. This has been going on since 1953 until today. The 200+ pages official English translated version of the 13th FYP (2016 – 2020) is divided into twenty parts (I – XX) and in eighty chapters, and covers miscellaneous topics, which are the scope for the current strategic evolvement of PRC (PRC13FYP, 2016). Reading through all pages is a lot to digest, but to conclude essential parts of the core; the overall message is to make technology advances in various prioritised scientific and technology research fields. Parts, or boxes, that might be of relevance with regards to cyber operations, are briefly described below and will later on be exemplified.

Box 3, Programs for Sci-Tech Innovation 2030 stipulates Science and technology programs with aim for aircraft engines and gas turbines, deep-sea stations, quantum communication and computing, brain science and brain-inspired research (a.k.a. artificial intelligence, neural networks), cyberspace security, deep space explorations and in-orbit spacecraft servicing and maintenance systems. It also mentions projects related to e.g. space-terrestrial information networks, big data, smart manufacturing and material science.

Box 7, High-End Equipment Innovation and Development aims for aerospace and marine engineering, transportation, high-level mechanical machine tooling, robotics, medical equipment and improved chemical manufacturing.

Box 8, Development of Strategic Emerging Industries stresses the importance of next generation of information technology industries – such as the 5G mobile communications, biotech, their Global Navigation Satellite System (GNSS) BeiDou, energy production and distribution, advanced and new materials such as next generation of semiconductor materials, and also so called new-energy vehicles as in all-electric and hybrid electric vehicles.

Box 9, Information Technology Projects has its focus on digital information, Internet and telecommunication expansions, Internet of Things (IoT), cloud computing innovation and development, the “Internet +” concept that aims for miscellaneous services where Internet is an integrated part, big data applications, better governmental e-services, e-commerce and lastly cybersecurity.

### **2.2 Made in China 2025 (中国制造2025)**

The national strategic plan Made in China 2025 (PRC, 2015) was launched in 2015 by Li Keqiang, the Premier of the State Council of PRC. The main goal with this plan is to strengthen the economic development. This will be done by improvements of industrialization and by year 2049 be the leading country in advanced green manufacturing. The program also aims to further develop information technology, the domains aerospace, aeronautics and oceanographic and underwater technology, rail transportation and automotive engineering, power production, agricultural machinery, material science, bio-pharmaceuticals and medical development. The MIC2025 appears to be highly connected with the 13th FYP.

## **3. Chinese information warfare and the evolvement of cyber doctrine**

It is besides the FYP and MIC2025 also worth to mention the military aspect of possessing intelligence information. The importance of information warfare and intelligence regarding military operations during the armed conflicts in e.g. the Middle East and Kosovo during 1990s and 2000s was noted by Chinese military strategists and later became a seed for the evolvement of their military cyber doctrine, where cyberspace was set to be a new security domain to defend (Jinghua, 2019).

## **4. Chinese offensive cyberspace operations**

Among the public resources at the cybersecurity company FireEye's website is their summary of identified advanced persistent threat (APT) groups (FireEye, 2020), also known as hacker groups. The listing briefly describes the suspected attribution country, targeted sectors, persona of the current group and alias if this is

known, malware that is associated with the attack and identified attack vectors. An extensive analysis report, blogs or even webinars can be associated with a reported APT. Six countries are mentioned by their names: Iran, PRC, North Korea, Russia and Vietnam. The numerous PRC related APTs are mainly targeting the sectors are depicted in Table 1.

**Table 1:** PRC-related APTs targeting various sectors

Sector	Sector (cont.)
Advertising	Information technology
Aerospace	International organisations
Agriculture	Investment
Automotive	Journalists
Aviation	Law firms
Biotechnology	Legal
Chemicals	Media
Construction and engineering/materials/manufacturing	Metals and mining
Defence	Military
Education	Mining
Electronics	Navigation
Energy	Non-profit organisations
Engineering	Pharmaceuticals
Entertainment	Public administration
Healthcare	Satellites and telecommunications
Financial services	Scientific research
Government	Telecommunications
Industrial engineering	Transportation

The various targeted industries are likely not to be a co-incidence. It aligns well with the strategic aims given in 13th FYP and MIC2025. The typical claimed outcomes of the attacks are exploiting of the targeted host, installed backdoors, exfiltration of stolen data and intellectual property (IP). What the cyber activities seem to have in common is the espionage perspective and at the same time keeping compromised servers alive.

The wide range of used attack vectors ranges from simple phishing to targeted spear-phishing emails with malicious payload, backdoor installations on compromised targets. The use of exploits, captured hosts to reach other targets, trusted relationship between targeted companies, command-and-control (C2, also CnC, C&C) servers, macro-enabled Microsoft Excel documents, watering hole attacks or strategic web compromises (SWCs), webshells and tools for crossing air-gapped networks are other examples of different modus operandi for succeeded APT campaigns. Extended information regarding APTs and their use of malware is more extensively described by the cybersecurity company The MITRE Corporation (MITRE, 2011). Research for designing attack infrastructure for offensive cyberspace operations also exists (Huskaj, Iftimie & Wilson, 2020).

#### 4.1 Operation Cloud Hopper

The main objectives for the notable Operation Cloud Hopper (also Cloudbopper) supply chain attack were cyberespionage, theft of IP and miscellaneous data of interest from the Microsoft Windows hosted managed service providers (MSPs) and their clients within the MSP’s cloud solution. The first intrusions against cloud-based MSPs are reported to be dated back to the beginning of 2016. It is claimed that the identified attributor APT10 (a.k.a “menuPass”, “Stone Panda”, “Red Apollo”, “CVNX” and “POTASSIUM”) at the time being were directed by the Ministry of State Security (MSS) via facilitating covered companies. The reported targeted MSPs are listed in Table 2:

**Table 2:** Targeted MSPs by Chinese cyberspace operations

MSP	MSPs (cont.)
Computer Sciences Corporation	International Business Machines (IBM) Corp
Dimension Data	NTT Data
DXC Technology	Tata Consultancy Services
Fujitsu	Visma
Hewlett Packard Enterprise (HPE)	

#### *4.1.1 Reported Modus Operandi*

Attack vectors that were used was a combination of spear-phishing; emails directed to employees containing executable malicious attachment in order to retrieve server access credentials, network reconnaissance, gain of elevated account credentials, use of various tailor-made malware for different services e.g. encrypted communication with C2 servers. The analysis of the Visma case showed that an APT10 installed version of WinRAR was utilised for compression of stolen data of interest, where data exfiltration were destined to Dropbox (PwC UK & BAE Systems, 2017; Bing, Stubbs & Menn, 2018; Bing, Stubbs & Menn, 2019; RecordedFuture, 2019; intrusiontruth, 2018; Kozy, 2018; Council of Foreign Relations, 2018; CISA, 2017).

#### *4.1.2 The aftermath of Operation Cloud Hopper*

On the 20th of December 2018, the United States Department of Justice makes a charge against the claimed MSS associated APT10 members Zhu Hua and Zhang Shilong for global computer intrusion campaigns targeting IP and confidential business information. The charges do not mention the Operation Cloud Hopper by its name, but indirectly suggesting them by mentioning two campaigns; The MSP Theft Campaign and The Technology Theft Campaign (DoJ, 2018b).

On the 30th of July 2020, the Council of the European Union (EU) imposes the first ever sanctions against cyberattacks in the sense of travel ban and freezing of possible assets, and in addition prohibition for EU citizens or entities to raise funds. For the Operation Cloud Hopper, the APT10 members Gao Qiang and Zhang Shilong are appointed as well as the company Tianjin Huaying Haitai Science and Technology Development Co. Ltd., which at the time being was suspected to be directed by MSS (European Council, 2020).

## **4.2 Other notable cyberattacks**

There are a numerous reported PRC associated cyberattacks of various types. Some of them directly or indirectly reflect the strived goals in 13th FYP. Others are meant to harm or demonstrate possessed cyberattack capabilities.

#### *4.2.1 India and the 40 000 cyberattacks attempts on banking and IT*

The low-intense 150 year old geopolitical Galwan Valley border conflict between PRC and India that once again begun to flourish on the 15th of June 2020, inflicted massive cyberattacks such as DDoS attacks, BGP hijacking, phishing emails and malware distribution against the India's IT infrastructure and banking sector, attributed by the APT3 (Gothic Panda) and APT10 (Stone Panda). This demonstrates some of the offensive capabilities (CYFIRMA, 2020; NDTV, 2020; India Today, 2020; Saha, 2020).

#### *4.2.2 Taiwan's semiconductor industry*

Box 8 in 13th FYP declares a strong interest in new semiconductor materials and related technology. One of the world's semiconductor centres is located in Taiwan and home for e.g. one of the industry leading company Taiwan Semiconductor Manufacturing Company (TSMC). CyCraft Technologies reported a series of new Skeleton Key attacks against the Taiwan semiconductor industry in the Operation Skeleton Key from late 2018 to the end of 2019 in an extensive analysis of Chimera APT group and the modus operandi. The claimed main objective for the attack was the theft of IP (CyCraft, 2020a; CyCraft, 2020b).

#### *4.2.3 The Comac C919 twinjet airliner*

Box 3 and Box 7 in the 13th FYP and MIC2025 denote the willingness in advances in aerospace technology and engineering. The Commercial Aircraft Corporation of China did almost succeed to develop a domestically designed airliner Comac C919. The development project was just need of a little help from the Jiangsu Province Ministry of State Security (JMSS) and MSS supported APT26 (Turbine Panda) between 2010 and 2015. A more thoroughly analysis of gathering trade secrets and IP is described in the report by CrowdStrike. Members of APT26 eventually led to a U.S. Department of Justice indictment in late 2018 (CrowdStrike, 2019; ThaiCert, 2020; Hruska, 2019; DoJ, 2018a).

#### 4.2.4 *The rumour of implanting surveillance devices*

The 4th of October 2018 did Bloomberg Businessweek publish the article *The Big Hack: How China Used a Tiny Chip to Infiltrate U.S. Companies*, reported that a security testing company discovered unintended tiny microchip on motherboards from the solution provider Super Micro Computer, Inc., or Supermicro. According to the article, one or several unnamed investigators concluded that the microchip would grant network access and claimed that the chips were post-installed on the motherboards at Supermicro subcontractor's factories "by operatives from a unit of the People's Liberation Army". As a consequence of the published article, Supermicro was shortly after delisted from NASDAQ and relisted in January 2020.

The claims of the spy chips and its functionality were later questioned by several other sources, including the U.S. Department of Homeland Security. The rumours did not only damage Supermicro. It also raised the question of subcontractors and secure supply chain logistics. More importantly, the awareness has risen regarding potential malicious design embedded into electronic designs (Robertson & Riley, 2018; DoH, 2018; Lee & Moltke, 2019; Baddeley, 2019; Greenberg, 2019; Hayes, 2020; Targett, 2020).

### 5. Sweden – the general situation

There is a wide range of published reports and surveys from Swedish authorities and from the private sector, where each publication concludes their version of the Swedish situation with regards to cyber operations, cyber incidents, intelligence, corporation damage, et cetera. Examples on such publications in random order are Englund (2019); Hanson et al. (2015); Swedish Security Service (2019); Bundgaard & Graflund-Wallentin (2020); Kristiansson (2019); ProofPoint (2020).

There is no official statistics or reliable figures on the true number of cyber operations that Sweden as a country has encountered, but it is reportedly at a constantly high number, and increasing every year (Olsson, 2019).

#### 5.1 Operation Cloud Hopper

It is claimed that the telecommunications company Ericsson, along with the ball bearings and rolling bearings company SKF, were two of the victims in the Operation Cloud Hopper as customers at the HPE breach. According to a spokesperson at SKF, their internal investigation concluded that no commercially sensitive data was stolen, while Ericsson made it clear that no customers were harmed (Bing, Stubbs & Menn, 2019).

The core businesses are of the two companies fits well into the plans of 13th FYP and MIC2025, where the SKF is continuously doing research for improvements of their products. Ericsson is considered to be a major 5G competitor to both Huawei and ZTE.

#### 5.2 The Swedish Protective Security Act and the 5G network infrastructure

From 1st of April 2019 did the new protective security legislation Protective Security Act (SFS, 2018a) and Protective Security Ordinance (SFS, 2018b) effect, as a consequence of the recent years' cyber operations as in cyberespionage and traditional espionage activates against vital targets in Sweden. The new legislation was proposed by the Swedish Government, where the main focus was to modernise and strengthen the security legislation on espionage, sabotage, terrorist acts and other hostile activities. Table 3 presents the statement from the Director General of the Swedish Security Service. Additional laws and information has been produced by the Swedish Security Service (2020a), Johansson (2018), and SFS (2018a & 2018b).

**Table 3:** Statement from the Director General of the Swedish Security Service

**Statement from the Director General of the Swedish Security Service**

*– China is one of the biggest threats to Sweden. The Chinese state is conducting cyber espionage to promote its own economic development and develop its military capabilities. This is done through extensive intelligence gathering and theft of technology, research and development. This is what we must consider when building the 5G network of the future. We cannot compromise with Sweden's security, says Klas Friberg.*

(Swedish Security Service, 2020b)

The Director General and Head of the Swedish Security Service (Säkerhetspolisen, Säpo) Mr. Klas Friberg expresses his concerns regarding PRC's cyberespionage and the expansion of the Swedish 5G network infrastructure in the matter of that China is considered one of Sweden's biggest threats, where the PRC conducts

cyberespionage campaigns for economic and military gains by intelligence and theft within the fields of technology, research and development – which Sweden has to consider during its 5G network expansion (Swedish Security Service, 2020b).

### **5.3 Military infiltration on the academia**

It is, as a parallel to cyberespionage, worth the mentioning of the academic perspective. Sweden – with its open universities, comparably low tuition fees, and an advanced level in science and technology research – offers higher levels of education to foreign students, as in master's or doctoral programmes. These kinds of opportunities have attracted students also from PRC. Among them are students sponsored by the People's Liberation Army's (PLA). The researcher Alex Joske at the Australian Strategic Policy Institute (ASPI) International Cyber Policy Centre published the report *Picking flowers, making honey – The Chinese military's collaboration with foreign universities* (Joske, 2018) where he describes the PLA systematic doctrine of sending and funding military scientists and engineers to abroad universities to study and to start collaborations with acknowledged western universities; technology transfer and intellectual property is unsuspectingly carried out from the Sci-Tech institutes to support research and development of domestic military projects. The report also mentions espionage and IP theft as additional duties for some of the students that are sent overseas. The report suggests several recommendations for restricting such activities. The military infiltration phenomenon is known to Swedish universities and university colleges (*högskolor*). It is however difficult to reveal such activities – and if known, hard to prevent. There is an information exchange on regular basis between universities and Säpo regarding the foreign students (Olsson, 2019; Eiderbrant, Sehlin & Johansson, 2019).

## **6. Conclusions**

The answer to the research question of 'how Chinese strategies are manifested into offensive cyberspace operations targeting Sweden' is noted not only in the Operation CloudHopper case, but also in military infiltration of the academia. Therefore, it is concluded that the Chinese have a three-pronged approach targeting the policy-level, industry and academia, using two disciplines: offensive cyberspace and espionage operations and human intelligence operations. The implications of this is an existential threat to Sweden's economic security and national security

It is obvious that People's Republic of China's (PRC) 13th Five-year plan and the additional Made in China 2025 strategy document have set the primary goal to raise the overall scientific and technology to a global top tier level. It can only be speculated in the reasons for this; political, military, social, scientific and technology influence, along total self-sufficiency in all kinds of manufacturing and production, from agriculture to green energy – but ultimately, an economic independence from other nations and to regain sovereignty.

Various cyber operations have proven to be vital tools for taking steps towards the defined goals in the political strategy. Several OSINT reports indicate that cyberespionage is one of the most common types of cyber operations, but it is likely to assume that PRC has the potential and willingness of issuing aggressive cyberwarfare campaigns – as in the Galwan Valley border conflict between India and China in June 2020.

From a Swedish perspective and with regards to the current political and trade relations with PRC, conduction of cyberattacks with the intention to bring down or disrupt the infrastructure, financial mechanisms or in any other way harm Swedish interests is here considered to be of lesser likelihood.

It is more probable that Sweden will continue to experience cyberespionage from PRC on all levels and on all domains; science, technology, IP and privacy information theft. Previously unveiled cyber operations exemplified in the essay have proven to be a convenient strategy for PRC to reduce its research and development gap in several ways; innovatively, financially and to shortening the time-to-market (TTM).

As a consequence of the Protective Security Act (SFS, 2018a) that went into force on 1st of January 2020, only four telecommunication companies were approved for the participation of the 5G frequency auctions arranged by the Swedish Post and Telecom Authority, Post- och telestyrelsen (PTS). Briefly, the licence conditions statutes that the licence holder shall safeguard that the use of the licensed radio frequency causes no harm to Sweden's security, installation of new radio infrastructure equipment from Huawei or ZTE is forbidden, existing equipment from these providers must be replaced before the 1st of January 2025, and staff or functions abroad must be relocated to Sweden before the same date (SFS, 2018a; The Swedish Post and Telecom Authority, 2020).



Allowing network and telecommunication equipment from Huawei or ZTE would potentially endanger the critical infrastructure for two major reasons: The hardware design, manufacturing and distribution of vital network components within PRC will always be questioned, similar to the rumour cause by Bloomberg Businessweek (Robertson & Riley, 2018); potential embedded espionage functionality will likely not be revealed if this is etched on the silicon die. The second and more worrying reason for the assumable endangerment is the National Intelligence Law of the People's Republic of China (2018 Amendment), or NIL, which was effective from the 27th of June 2018. Article 7 reads as follows (English translated version) (LawinfoChina.com, 2018). The applicability of NIL is further analysed in the report by Dackö & Jonsson (2019), where a previous version of NIL effective on 27 June 2017 is legally interpreted.

**Table 4:** Article 7 of the National Intelligence Law of the PRC

<p><b>Article 7 of the National Intelligence Law of the PRC.</b></p> <p><i>An organization or citizen shall support, assist in and cooperate in national intelligence work in accordance with the law and keep confidential the national intelligence work that it or he knows.</i></p> <p><i>The state shall protect the individual or organization that has supported, assisted in or cooperated in national intelligence work.</i></p> <p>(LawinfoChina.com, 2018)</p>
---

The Picking Flowers – Making Honey report (Joske, 2018) highlights a number of potential risks of having foreign students in the Swedish universities. This would assumingly have a greater negative impact on a doctoral level, where availability of immaterial property and unpublished research data is at hand. This sort of intelligence information harvesting applies well to the many of the goals in 13th FYP. However, most of the foreign students are statistically not funded by the PLA. The benefits of having foreign students from PRC outweigh the risks in many ways. The international environment and the chance of growing global contact networks are more valuable.

## 6.1 Future work

Future work can study the financial impact of Chinese offensive cyberspace and espionage operations combined with human intelligence operations.

## Appendix 1: Acronyms

Acronym	Meaning
13th FYP	13th Five-years plan
APT	Advanced Persistent Threat
DDoS	Distributed Denial-of-Service
IP	Intellectual property
OSINT	Open-source intelligence
MIC2025	Made in China 2025
MSP	Managed service provider
MSS	Ministry of State Security
NIL	National Intelligence Law of the People's Republic of China
PLA	People's Liberation Army
PRC	People's Republic of China, in this context the government and military
Säpo	Säkerhetspolisen, Swedish Security Service
PTS	Post- och telestyrelsen, Swedish Post and Telecom Authority

## References

- Baddeley, B. (2019). "What Happened With Supermicro?" Retrieved from <https://hackaday.com/2019/05/14/what-happened-with-supermicro/>.
- Bing, C., Stubbs, J., Menn, J. (2018). "Exclusive: China hacked HPE, IBM and then attacked clients – sources." Retrieved from <https://www.reuters.com/article/us-china-cyber-hpe-ibm-exclusive/exclusive-china-hacked-hpe-ibm-and-then-attacked-clients-sources-idUSKCN1OJ2OY>.
- Bing, C., Stubbs, J., Menn, J. (2019). "Inside the West's failed fight against China's 'Cloud Hopper' hackers." <https://www.reuters.com/investigates/special-report/china-cyber-cloudhopper/>.
- Bundgaard, J. & Graflund-Wallentin, S. (2020). "Nordic Cyber Crime Survey 2020: Det digitaliserade Sverige – så in i Norden säkert? En undersökning om cybersäkerheten i Sverige, Norge och Danmark." Retrieved from <https://www.pwc.se/sv/cyber-security/cyberbrottslighet.html>.

- CISA. (2017). "Intrusion Affecting Multiple Victims Across Multiple Sectors." Retrieved from <https://us-cert.cisa.gov/ncas/alerts/TA17-117A>.
- Council of Foreign Relations. (2018). "Compromise of Managed Service Providers and technology companies." Retrieved from <https://www.cfr.org/cyber-operations/compromise-managed-service-providers-and-technology-companies>.
- CrowdStrike. (2019). "Huge Fan of Your Work: How TURBINE PANDA and China's Top Spies Enabled Beijing to Cut Corners on the C919 Passenger Jet." Retrieved from <https://www.scribd.com/document/430534695/Crowdstrike-Huge-Fan-of-Your-Work-Intelligence-Report>.
- CyCraft Research Team. (2020a). "Craft for Resilience, APT Group Chimera – APT Operation Skeleton Key Targets Taiwan Semiconductor Vendors." Retrieved from [https://cycraft.com/download/%5BTLP-White%5D20200415%20Chimera\\_V4.1.pdf](https://cycraft.com/download/%5BTLP-White%5D20200415%20Chimera_V4.1.pdf).
- CyCraft Technology Corp. (2020b) "Taiwan High-Tech Ecosystem Targeted by Foreign APT Group: Digital Skeleton Key Bypasses Security Measures." Retrieved from <https://medium.com/cycraft/taiwan-high-tech-ecosystem-targeted-by-foreign-apt-group-5473d2ad8730>.
- CYFIRMA. (2020). "Rising cyber attacks due to China-India border conflict." Retrieved from <https://www.cyfirma.com/early-warning/rising-cyber-attacks-due-to-china-india-border-conflict/>.
- Dackö, C. & Jonsson, L. (2019). "Applicability of Chinese National Intelligence Law to Chinese and non-Chinese Entities." Retrieved from [https://www.mannheimerswartling.se/globalassets/nyhetsbrev/msa\\_nyhetsbrev\\_national-intelligence-law\\_jan-19.pdf](https://www.mannheimerswartling.se/globalassets/nyhetsbrev/msa_nyhetsbrev_national-intelligence-law_jan-19.pdf).
- Eiderbrant, A., Sehlin, A., & Johansson, L. "Rekordmånga kinesiska studenter i Stockholm – KTH har kontakt med Säpo." Retrieved from <https://www.svt.se/nyheter/lokalt/stockholm/kinesiska-studenter>.
- Englund, J. (2019). "Kinas industriella cyberspionage." Retrieved from <https://www.foi.se/rapportsammanfattning?reportNo=FOI%20MEMO%206698>.
- European Council. (2020). "EU imposes the first ever sanctions against cyber-attacks." Retrieved from <https://www.consilium.europa.eu/en/press/press-releases/2020/07/30/eu-imposes-the-first-ever-sanctions-against-cyber-attacks/>.
- FireEye (2020) "Advanced Persistent Threat Groups – Who's who of cyber threat actors." Retrieved from <https://www.fireeye.com/current-threats/apt-groups.html>.
- Greenberg, A. (2019) "Planting Tiny Spy Chips in Hardware Can Cost as Little as \$200." Retrieved from <https://www.wired.com/story/plant-spy-chips-hardware-supermicro-cheap-proof-of-concept/>.
- Hanson, M., Johansson, T., Lindgren, C. & Oehme, R. (2015). "MSB851, Information Security – trends 2015, A Swedish perspective", Retrieved from <https://www.msb.se/siteassets/dokument/publikationer/english-publications/information-security--trends-2015-a-swedish-perspective.pdf>.
- Harknett, R.J., & Smeets, M. (2020) Cyber campaigns and strategic outcomes, Journal of Strategic Studies, DOI: 10.1080/01402390.2020.1732354.
- Hayes, P.G. (2020). "Supermicro® Announces Approval to Relist on NASDAQ and Provides Business Update." Retrieved from <https://www.supermicro.com/en/pressreleases/supermicro-announces-approval-relist-nasdaq-and-provides-business-update>.
- Hruska, J. (2019). "Report: China's New Comac C919 Jetliner Is Built With Stolen Technology." Retrieved from <https://www.extremetech.com/extreme/300313-report-chinas-new-comac-c919-jetliner-is-built-with-stolen-technology>.
- Huskaj, G., Iftimie, I.A. & Wilson, R.L. (2020). Designing Attack Infrastructure for Offensive Cyberspace Operations. In: Proceedings of the 19th European Conference on Cyber Warfare and Security: a virtual conference hosted by University of Chester UK 25-26 June 2020 / [ed] Thaddeus Eze, Lee Speakman, Cyril Onwubiko, Reading, UK: Academic Conferences and Publishing International Limited, 2020, pp. 473-482.
- India Today. (2020). "Cyber Attack: Chinese Hackers Attempted 40,000 Cyber Attacks On Indian Web, Banking Sector In 5 Days." Retrieved from <https://www.indiatoday.in/india/story/chinese-hackers-attempted-40-000-cyber-attacks-on-india-1692088-2020-06-24>.
- intrusiontruth. (2018). "APT10 was managed by the Tianjin bureau of the Chinese Ministry of State Security." Retrieved from <https://intrusiontruth.wordpress.com/2018/08/15/apt10-was-managed-by-the-tianjin-bureau-of-the-chinese-ministry-of-state-security/>.
- Jinghua, L. (2019). "What are China's Cyber Capabilities and Intentions?", Retrieved from <https://carnegieendowment.org/2019/04/01/what-are-china-s-cyber-capabilities-and-intentions-pub-78734>.
- Johansson, M. (2018). "Regeringens proposition 2017/18:89, Ett modernt och stärkt skydd för Sveriges säkerhet – ny säkerhetsskyddslag." Retrieved from <https://www.regeringen.se/rattsliga-dokument/proposition/2018/02/prop.-20171889/>.
- Joske, A. (2018). "ASPI International Cyber Policy Centre, Policy Brief Report No. 10/2018, Picking flowers, making honey – The Chinese military's collaboration with foreign universities." Retrieved from <https://www.aspi.org.au/report/picking-flowers-making-honey>.
- Kozy, A. (2018). "Two Birds, One STONE PANDA." Retrieved from <https://www.crowdstrike.com/blog/two-birds-one-stone-panda/>.
- Kristiansson, S. (2019). "Underrättelsehotet mot Sverige." Retrieved from <https://frivarld.se/rapporter/underrattelsehotet-mot-sverige/>.

- LawInfoChina.com. (2018). Standing Committee of the National People's Congress (27 April 2018), "National Intelligence Law of the People's Republic of China (2018 Amendment)." Retrieved from <http://www.lawinfochina.com/display.aspx?id=28135&lib=law>.
- Lee, M. & Moltke, H. (2019). "Everybody Does It: The Messy Truth About Infiltrating Computer Supply Chains." Retrieved from <https://theintercept.com/2019/01/24/computer-supply-chain-attacks/>.
- NDTV. (2020). "Rise In Cyber Attacks From China, Over 40,000 Cases In 5 Days: Official." Retrieved from <https://www.ndtv.com/india-news/rise-in-cyber-attacks-from-china-over-40-000-cases-in-5-days-official-2251111>.
- Olsson, J. (2019). "De forskar för Kinas militär." Retrieved from <https://universitetslararen.se/2019/02/13/de-forskar-for-kinas-militar/>.
- Olsson, J. (2019). "FRA: Cyberangrepp mot Sverige ökar." Retrieved from <https://www.svt.se/nyheter/fra-cyberangreppen-mot-sverige-okar>.
- PRC. (2015). People's Republic of China, State Council (7 July 2015), "Made in China 2025, 中国制造2025." Retrieved from <http://www.cittadellascienza.it/cina/wp-content/uploads/2017/02/loT-ONE-Made-in-China-2025.pdf>.
- PRC. (2016). Central Committee of the Communist Party of China, translated by Compilation and Translation Bureau, "The 13th five-year plan for economic and social development of The People's Republic of China 2016-2020." Retrieved from [https://www.un-page.org/files/public/china\\_five\\_year\\_plan.pdf](https://www.un-page.org/files/public/china_five_year_plan.pdf).
- Proofpoint Inc. (2020). "People-Centric Cybersecurity: A Study of IT Security Leaders in Sweden." Retrieved from [https://www.proofpoint.com/sites/default/files/2020-06/Proofpoint\\_2020\\_SW\\_CISO\\_SURVEY\\_REPORT\\_Final\\_0.pdf](https://www.proofpoint.com/sites/default/files/2020-06/Proofpoint_2020_SW_CISO_SURVEY_REPORT_Final_0.pdf).
- PwC UK and BAE Systems. (2017). "Operation Cloud Hopper." Retrieved from <https://www.pwc.co.uk/cyber-security/pdf/cloud-hopper-report-final-v4.pdf>.
- RecordedFuture. (2019). "APT10 Targeted Norwegian MSP and US Companies in Sustained Campaign." Retrieved from <https://www.recordedfuture.com/apt10-cyberespionage-campaign/>.
- Robertson, J., & Riley, M. (2018). "The Big Hack: How China Used a Tiny Chip to Infiltrate U.S. Companies." Retrieved from <https://www.bloomberg.com/news/features/2018-10-04/the-big-hack-how-china-used-a-tiny-chip-to-infiltrate-america-s-top-companies>.
- Saha, A. (2020). "40,000 cyber-attacks attempted by Chinese hackers on Indian banking, IT sector in five days." Retrieved from <https://www.dnaindia.com/india/report-40000-cyber-attacks-attempted-by-chinese-hackers-on-indian-banking-it-sector-in-five-days-2829381>.
- SFS. (2018a). "SFS 2018:585, Säkerhetsskyddslag." Retrieved from <https://svenskfattningssamling.se/doc/2018585.html>.
- SFS. (2018b). "SFS 2018:658, Säkerhetsskyddsförordning." Retrieved from <https://svenskfattningssamling.se/doc/2018658.html>.
- Swedish Security Service. (2019). "Årsbok 2019." Retrieved from <https://www.sakerhetspolisen.se/publikationer/om-sakerhetspolisen/sakerhetspolisen-2019.html>.
- Swedish Security Service. (2020a). Protective security. Retrieved from <https://www.sakerhetspolisen.se/en/swedish-security-service/protective-security.html>.
- Swedish Security Service. (2020b). "Säkert 5G viktigt för Sverige." Retrieved from <https://www.sakerhetspolisen.se/ovrigt/pressrum/aktuellt/aktuellt/2020-10-20-sakert-5g-viktigt-for-sverige.html>.
- Targett, E. (2020). "What Supermicro Did Next." Retrieved from <https://www.cbronline.com/interview/what-supermicro-did-next>.
- ThaiCERT. (2020). "Threat Group Cards: A Threat Actor Encyclopedia." Retrieved from <https://apt.thaicert.or.th/cgi-bin/showcard.cgi?g=Turbine%20Panda%2C%20APT%2026%2C%20Shell%20Crew%2C%20WebMasters%2C%20KungFu%20Kittens&n=1>.
- The MITRE Corporation (2011). MITRE ATT&CK, "Groups." Retrieved from <https://attack.mitre.org/groups/>.
- The Swedish Post and Telecom Authority. (2020). "Assignment in the 3.5 GHz and 2.3 GHz bands." Retrieved from <https://www.pts.se/en/english-b/radio/auctions/assignment-in-the-3.4---3.8-ghz-bandet/>.
- The United States Department of Justice. (2018a). "Chinese Intelligence Officers and Their Recruited Hackers and Insiders Conspired to Steal Sensitive Commercial Aviation and Technological Data for Years." Retrieved from <https://www.justice.gov/opa/pr/chinese-intelligence-officers-and-their-recruited-hackers-and-insiders-conspired-steal>.
- The United States Department of Justice. (2018b). "Two Chinese Hackers Associated With the Ministry of State Security Charged with Global Computer Intrusion Campaigns Targeting Intellectual Property and Confidential Business Information." Retrieved from <https://www.justice.gov/opa/pr/two-chinese-hackers-associated-ministry-state-security-charged-global-computer-intrusion>.
- Thornton-Trump, I. (2019) THE POLITICS OF CYBER, EDPACS, 59:3, 1-17, DOI: 10.1080/07366981.2019.1564193
- U.S. Department of Homeland Security. (2018). "Statement from DHS Press Secretary on Recent Media Reports of Potential Supply Chain Compromise." Retrieved from <https://www.dhs.gov/news/2018/10/06/statement-dhs-press-secretary-recent-media-reports-potential-supply-chain-compromise>.

# The Evolution of Cyber Fraud in the Past Decade

George-Daniel Bobric

“Carol I” National Defense University, Bucharest, Romania

[dbobric08@gmail.com](mailto:dbobric08@gmail.com)

DOI: 10.34190/EWS.21.010

**Abstract:** Everyday reality faces an existential paradigm: the belligerent and puritanical visions characteristic to the beginning of the current century were projected from the physical space into the operational environment constituted by the cyberspace. Similarly, cyberspace is one of the “grounds” used for initiating illicit operations carried out by various individuals or groups to achieve personal or collective goals. The peculiarities of the cyber environment that favour the commission of online crimes under the auspices of the significant protection of the perpetrator’s real identity materialize in a unitary whole that, in recent years, has been irrefutably consolidated and which represents a major threat to the states’ national security. At the same time, the concoction of motives that constitute the starting point in the process of elaborating, initiating and executing cyber attacks revolves around ensuring the personal gains of their initiators, especially the financial ones. This paper is an empirical, qualitative research, its objective being to investigate the evolution of the actions performed within the cyberspace related to the cyber-enabled fraud, both in terms of cantitative and qualitative aspects. In order to achieve the proposed objective, an analysis of the literature relevant to the topic of the paper was performed. Also, significant data provided by institutions from different countries with a special role in the cyber security domain were collected, to have an overview on the current situation regarding the cyber fraud activities by analyzing the trends from the past decade. Nonetheless, a short presentation of the possible evolutive directions of the instruments specific to the cyber fraud in the future years will be performed. The results of the study show a vertiginous increase from year to year in the number of cyber-fraud actions, in the number of the victims’ complaints and in the amount of financial losses registered as a result of these actions. Starting from these preliminary data, the profile organizations can elaborate future in-depth studies on this worrying phenomenon represented by the exacerbation of the number of illicit actions carried out in the cyberspace categorized as cyber-fraud.

**Keywords:** cyber-fraud, cyberspace, cyber attacks, cybercrime, cybersecurity

---

## 1. Introduction

### 1.1 General background

In a world where the national security is the sum of factors characterized by a constant state of volatility, the level of safety that hovers over the individual is directly dependent on the interactions with other environments: technological, informational, financial, digital etc. In this sense, there is a close interaction between national security and individual safety, generated by the functional and syntactic interdependence between the two words with seemingly identical meaning. In the case of a system that is part of an environment, there are two possibilities. Security is represented by the reduced possibility of the environment to critically affect the system, and security consists in the impossibility of the system to seriously affect the environment (Line et al., 2006). Therefore, there is an interaction between the two concepts which is often affected by a high number of external actions, coming from other environments.

One such environment is cyberspace, the area in which actors from various parts of the world act, for different reasons, with different goals to achieve, and in which space and time are constrained compared to the actions performed in the physical environment. In this way, the footprint that each cyber actor leaves as a result of his actions is diverse, and their impact may vary depending on the level of knowledge, objectives and technique available to the individual. The effects of these actions have, in a broad sense, consequences for national security leading to reverberations for the individual security of the population. Malicious actors have identified the malignant potential of cyberspace and the benefits that can be used to perpetuate or support the actions performed in the physical environment. A concrete example in this sense is the organized crime, individuals involved in carrying out these actions in the physical environment leading these activities to the virtual environment.

### 1.2 General purpose and structure of the paper

This study represents an analysis of the evolution of the cyber actions related to physical or electronic fraud activities, in terms of the number and typology, elaborated based on the information provided by some institutions with responsibilities in the field of cybersecurity.

Regarding the structure of the paper, after the introduction, a short theoretical foray into the field of cybercrime will be performed, presenting the main ideas related to the core of this paper. Subsequently, a main component that is part of the spectrum of activities included within the cybercrime register, cyber fraud, will be analyzed. Next, the paper continues with a brief presentation of the actors that use the cyberspace in order to defraud the victims, through the prism of two points of view: the motivation underlying the actions, respectively the social classification of individuals/groups. Subsequently, the paper continues with the analysis of the evolution of cyber fraud in the last decade, taking into account data from the United States and Europe. Finally, the paper presents possible further directions of action that could lead to new threats to personal security generated by the perpetuation of the actions specific to organized crime, the paper concluding with the presentation of the main conclusions that emerge from the analysis.

### **1.3 The research methodology**

The current paper is an analysis of the evolutionary trend of the number of actions performed in cyberspace related to cyber-fraud. The scientific research method that underlies this paper is the method of bibliographic documentation, used to identify information relevant to the proposed topic, collecting raw or analyzed data on cases of cyber fraud in recent years, enriching the knowledge substrate with current information, etc. At the same time, another research methods used are the analysis and comparison of the data provided by some institutions from United States of America and Europe with responsibilities in the field of cybersecurity.

### **1.4 Literature review**

In this section, a brief analysis of the previous work related to the subject of cyber-enabled fraud will be performed. Nurse (2018) presents an interdisciplinary analysis of the cybercrime that particularly target individuals, cyber fraud having a special role regardless the motivation behind the actions. A paper presented by Moneva (2020) shows that, in the previous year, cybercrime faced an increase in April and in the following months as a result of the lockdown related to the pandemic situation, whilst the cyber-enabled fraud showed a less variation, even if it is increasingly larger in number. Another relevant work for the current paper's theme is a report of the European Payments Council (2019) regarding the evolution and developments of the cyber fraud-related activities over the past years, presenting significant information about actors, instruments, types of criminal cyber attacks, virtual currencies and different types of fraud. Despite the other two mentioned papers, the last one is a comprehensive study aiming to provide an in-depth analysis of an evolving phenomenon, whilst the other two papers focus on particular aspects rather than on the process as a whole. Nonetheless, Button & Cross (2017) analyzed the interrelation between the evolution of technology in the past years and the computer-fraud, concluding with the idea that the technological changes generated important gaps in different societal areas, thus leading to different consequences such as those generated by the cybercrime, in a wider approach, and by the computer-fraud, in a smaller extent.

## **2. General overview of cybercrime**

### **2.1 What is cybercrime?**

Cybercrime is a domain that evolves along with the developments from cyberspace and its related infrastructure. Currently, crime in the virtual environment is not independent, but is part of a more complex whole, based on which there are achieved the personal or collective objectives of individuals who carry out such actions. The Internet, through its peculiarities, facilitated operations in the physical environment, not only in terms of facilitating the communication between the members of the same network, but especially in that it transposed methods and means of crime from the real world into the virtual world. At the same time, the Internet has led to the expansion of the range of actions specific to crime in the physical environment, as well as to the provision of new benefits for traditional operations, which through cyberspace have become more efficient, more profitable and much less risky.

In general, although there is currently no unanimously accepted definition, cybercrime can be presented as the set of illicit actions carried out by various cyber actors on the digital or information environment, in which the information system is either used as a tool or viewed as a target (Sabillon et al., 2016). In the specialty literature, four types of cybercrime have been identified, related to the relation between computer and action: computer as a target, computer as a tool, computer as an alternative resource and criminal activities associated with the

expansion of cybercrime. The first category focuses on information content stored on a computer, and the second category includes actions specific to cyber fraud (Jahankhani, Al-Nemrat, Hosseinian-Far, 2014).

## **2.2 Aspects related to cyber-enabled fraud**

According to the Cambridge Dictionary, the word “fraud” is the “crime of making money by deceiving people, a person who pretends to be who he is not in order to deceive people or something who is not what it appears to be and is used to deceive people, especially to obtain financial resources”. Extrapolating to the cybernetic field, cheating people through the virtual environment to obtain financial income represents the cyber-fraud.

One of the most common forms of online fraud is credit card fraud, which involves one of the methods used by cybercrime groups: identity theft. This is performed for withdrawing/transferring money from the victim's accounts, purchasing goods/services by paying their equivalent value with the victim's credit card or using her data to perpetuate other illegal activities. Increasing the intensity of the damage, the individual can use the victim's account to obtain loans (amounts of money that do not involve an additional contract with the bank for the current account) or to open new accounts. In interaction with these activities mentioned above is the technique called “salami fraud” or its equivalent “collect the roundoff”, through which the attacker constantly collects small amounts of money that do not generate the detection of this action. The “collect the roundoff” technique refers to the collection of monetary fractions generated by rounding certain financial calculations to the nearest whole number. Over time, these techniques of illicit collection of money from victims can generate substantial income for the initiator of the action (Alhassan, 2018). Another genuine example of cyber fraud consists of phishing, which is a very common method that involves sending e-mails to individuals who claim to be part of the team of well-known companies. These messages usually include advertisements for various goods or services preferred by users (electronic devices, clothing, etc.) and through which the idea of a potential gain is inoculated to the victim. By filling in the fields and transmitting the personal or bank accounts related information, the process returns to the two techniques above-mentioned: identity theft and credit card fraud (Danhioux, 2013).

Cyber fraud can take various forms, more and more diverse and sophisticated. One of the latest examples of cyber fraud is the case of two women in the UK who became victims of a scam on social media (\*\*\*, 2020). The first one met the individual who defrauded her through a dating site, asking her, over time, certain amounts of money for the business he was supposedly administering. The amount the scammer received from the victim amounted to a total of £320,000. For the victim to gain confidence, the scammer suggested that he sent the money to the translator, the second victim. The latter was convinced that she was in a relationship with the false identity of the scammer and that the amount of money received from the first woman was part of an investment scheme, and it had to be deposited in another account (of the scammer). The English police's investigation determined that the money transferred by the second woman was placed in a money laundering circle in the Far East. Fraud activities carried out by various organized crime organizations or by various individuals acting alone in cyberspace have their origin in the actions generally carried out in the physical environment. The hacking correspondent in the physical space is the unauthorized intrusion into the public or private space. Online data theft has the same equivalent in the real world. The seizure of goods or their theft is similar to cyber-specific denial-of-service attacks. Forging real-life documents lead to the same type of activity in the virtual environment. Actions that infringe copyright in real life have about the same meaning and impact as activities specific to intellectual property crime. In this idea, the actions of cyber fraud are not a revelation for criminals, but a method of supporting illicit actions carried out in the physical environment or multiplying the outcomes (financial, intellectual, etc.).

## **3. Actors perpetuating cyber-enabled fraud**

As for the components of cybercrime, they are diverse, taking into account the reasons behind such actions and the objectives that are to be achieved. Generically, groups that perform activities that can be included in the range of illicit actions specific to organized crime are divided into four main categories (Nurse, Bada, 2018): groups that use cyber facilities to support the physical ones, groups that perform actions especially in cyberspace, ideologically and/or politically motivated groups, respectively groups of citizens who mobilize through cyberspace and act. The first category includes those people who use cyberspace to facilitate transactions, to launder money, to negotiate/establish various details of future operations, etc. The second category includes individuals who operate specifically in cyberspace to perform various operations such as cyber fraud, extortion, blackmail, etc. The third category includes ideologically motivated people (both morally and

religiously), an example being the terrorists, as well as politically motivated people or those reacting to decisions taken by political decision-making bodies. Last but not least, the fourth category includes people who use the cyber environment as a tool to join the group and increase its cohesion, the group reacting to a certain issue at a certain time. Although the last two categories do not seem to be related to organized crime, the potential for these movements to transform into or generate illicit actions is high, or in actions that can degenerate into manifestations specific to organized crime.

According to another classification of actors who carry out operations in the range of cyber fraud, they can be classified into four main categories which, taking into account the criterion of general relevance, are: opportunistic cybercriminals, legitimate organizations, organized crime networks and foreign intelligence services. (Detica, 2011). Opportunistic cybercriminals are the lowest level of threat in terms of cyber fraud, given the fact that obtaining financial benefits is, in most cases, based on less sophisticated techniques and methods of action, and their impact on victims is mainly at a low level. This is due to the dependence of the level of sophistication of the actions undertaken on cyber training and the resources available to the perpetrator. Legitimate organizations can engage in cyber actions in their competition with others, and cyber fraud can be the mainstay of a wide range of actions to ensure competitive advantage in the fight for supremacy, based on the principle that the purpose excuses the means. Organized crime networks are the actors that operate frequently and dominantly in relation to other actors in the cyberspace due to the facilities offered by this operational field and its peculiarities that ensure high achievements with minimal investments and risks. At the same time, the cyber environment is the optimal space both for carrying out activities specific to organized crime (including cyber-enabled fraud) and for facilitating and improving the activity in the physical environment, respectively for maximizing profits. The highest level of action sophistication is that of foreign intelligence services. Although the direct involvement of them is performed only in certain situations, determined by the achievement of specific objectives at the strategic level through the use of tools specific to cybercrime, this typology can lead to more sophisticated and better-developed actions, having at their disposal a wide range of resources (financial, informational, material, etc.).

#### **4. The evolution of cyber fraud in the past years**

Over time, the actions by which the actors in the physical environment consolidate their position in different fields of activity are diversifying, the tools through which they are performed evolving unequivocally. This aspect also applies to the cyber environment, in view of the fact that individuals who carry out illegal activities in the real environment seek to diversify the mode of action by transposing them into cyberspace. Starting from this hypothesis, in this subchapter the evolution of actions that subscribe, generically, to cyber fraud, will be analyzed, by comparing the relevant data identified in the scientific reports of some structures with specific attributions in the cybersecurity domain.

From a quantitative point of view (with reference to the financial impact, on the one hand, and the number of actions performed, on the other hand), the situation is on an upward trend, registering an increase both in terms of financial losses as a result of these actions, as well in the number of actions. The annual report of the Federal Bureau of Investigation of the United States of America presents an increase in both the number of complaints against cyber fraud and the losses recorded by victims in the 2015-2019 period (Gorham, 2019). If in 2015 there were approximately 290,000 complaints in the United States, by 2019 their number increased by about 62 percent, reaching about 467,000 complaints. In the same vein, the financial losses generated by these actions increased by more than 300 percent over the same period (Figure 1). According to the same report, cyber fraud was ranked second in the top of criminal activities carried out in 2019. According to the aforementioned report, the losses in 2019 due to the use of the false technical support method increased by 40 percent compared to the previous year.

Regarding the number of actions carried out in the cyberspace that can be included in the sphere of cyber fraud, in order to identify the way in which it evolved, the data provided by certain CERT (Computer Emergency Response Teams), also known as CSIRT (Computer Security Incident Response Teams) - in this case Lithuanian (CERT-LT), Polish (CERT-PL), Romanian (CERT-RO) and Slovak (CSIRT-SK) - will be analyzed. The data are shown in Figure 2.

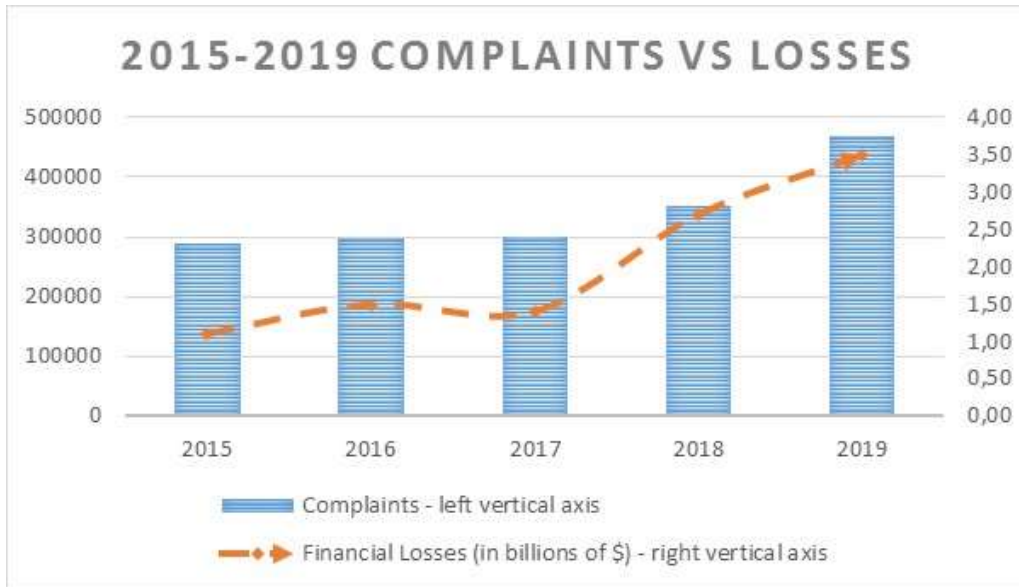


Figure 1: Complaints vs. loses

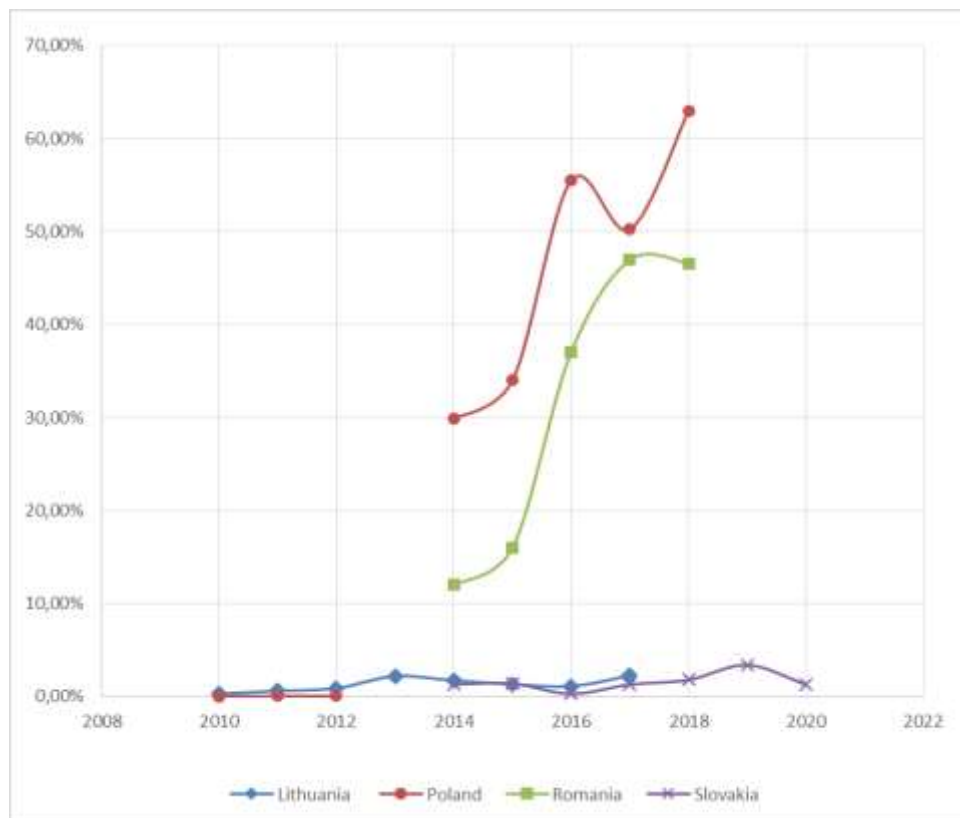
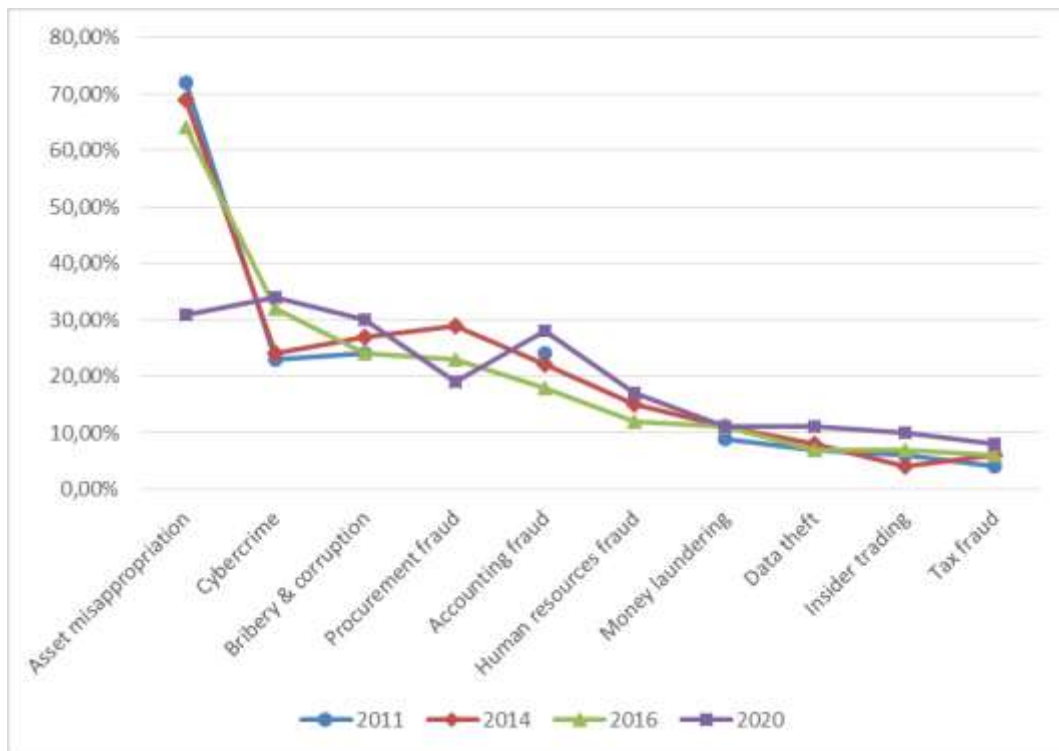


Figure 2: Variations of the number of attacks related to computer fraud

From figure 2 it can be seen that the number of cases of actions in cyberspace related to computer fraud has increased significantly, in 2017 about half of the manually processed registered attacks being actions from the online fraud category. In the second figure, the significant difference between the values is generated by the way of processing the attacks registered by each state during a calendar year, in this case the values in the range 1-3 percent being those processed automatically, and those in the range 10-65 percent being processed manually, having a much higher relevance and degree of correctness of the assignment. In this sense, one of the main limitations of this study identified is the lack of concrete data on the number of actions in cyberspace that can be related to computer fraud registered by different state entities during a year, whether they are processed manually or automatically.



The same trend can be identified by performing a combined qualitative and quantitative analysis, by dividing the computerized fraud into its component parts and analyzing the number of actions identified by the profile institutions for each subcomponent. This can be seen from Figure 3, which was realized based on the information provided by PwC reports.



**Figure 3:** Variation of the actions specific to the computer fraud's subdomains

In the last decade, the number of asset misappropriation actions has declined sharply, while the percentage of cybercrime-specific actions has grown at an almost constant rate. In this sense, with all the rules applied in the last decade, on the basis of an increase in the number of fraudulent cyber actions, it can be observed maintenance of an approximately constant trend of the other subdomains specific to computer fraud, with slight variations from case to case and from year to year.

## 5. Future possible directions in the field of cyber-enabled fraud

Criminal organizations have a wide range of tools at their disposal that can be used to achieve their own goals and, sometimes, those of third parties. Nation-states finance these activities to obtain benefits from the targets (especially their opposing states and important organizations) or to stimulate the maintenance of a certain level of instability of that entity. In this sense, foreign intelligence services have the necessary levers to fulfil the tasks set by the leadership of the state by using criminal organizations. In the future, these tools are expected to be used extensively, being a quick way to achieve the objectives.

Organized crime networks will most likely use the cyber environment to facilitate operations in the physical environment and to maximize personal gains while minimizing risks. At the same time, the innovations in the technological field related to cyberspace will be used by the members of these networks in order to adapt the way of carrying out the actions to the new inventions. An example of this is the use of artificial intelligence to perform vishing actions (social engineering using an automatic voice). Although this seems, at the moment, in the field of science fiction, in reality, this fact was first performed in 2019, when an organized crime group defrauded an energy company through this technology. The group replicated the voice of the company's CEO, initiated the social engineering operation on a person within the company who authorized the transfer of \$243,000 to a fake account due to the fact that he was deceived by the replicated voice of his boss. To give credibility to the action, the group used a fake profile of a company on social networks, with fake online content and fake people representing the alleged management team of the alleged company (Durbin, 2020).

Moreover, research and innovation in the field of information technology continue, and the development of this field in the coming years is inherent. Recently, about a year ago (October 2019), Google announced an innovation in information technology, in this case solving a problem that would have lasted 10,000 years for the most modern and high-performance computer today in just a few minutes with the new quantum computer (Reuters, 2019). Such an achievement leads us to the idea that the future of information processing is at a turning point, and the corroboration of this innovation with the discoveries in the field of artificial intelligence can lead to an era in which data processing will be performed with amazing speed. This can create a new set of threats to the personal security of the individual, including through the use, by criminal organizations, of these new technologies for the practice of obscure actions, including those specific to cyber-enabled fraud.

## **6. Conclusion**

The security situation deriving from the use of the cyber environment by different state or non-state entities evolves simultaneously with the emergence of new technologies that improve its particularities. Organized crime groups use cyberspace, on one hand, to facilitate illicit operations in real space, and on the other hand, to carry out actions specific to organized crime (such as fraud) in an environment that maximizes the chances of success at the same time as minimizing the related risks. As it can be concluded from the fourth part of the present paper, the threats deriving from the illicit use of the cyberspace by different actors in order to perform actions related to cyber-fraud are emerging. This idea occurs, as presented above, due to the following aspects:

- in the second part of the past decade, an increase in the number of complaints initiated by the victims of the actions related to cyber-enabled fraud has been recorded, as in the case of the United States analyzed above. Starting from this point, it can be stated that the same trend could have been seen in other parts of the world, as well as for longer periods of time.
- in the 2015-2019 period, the amount of financial losses increased by 4.5 times in the case of United States, according to the available data. Based on this fact, it seems plausible to advance the hypothesis that, in the last year, the value of financial losses also increased and, without any relevant actions taken against cyber fraud perpetrators, the value of financial losses will also increase in the following years.
- in the 2014-2018 period, the data regarding the actions related to cyber fraud manually analyzed by the Romanian and Polish authorities with responsibilities in the field of cyber security show an important increase in the percentage of actions specific to cyber fraud from the total number of actions monitored and analyzed.
- with respect to the categories of actions that fall under the “umbrella” of the cyber fraud, a downward trend of the asset misappropriation can be seen, showing that people are getting acquainted with the correct use of the assets related to the digital domain. Also, in the past decade, all the other categories of cyber fraud faced a slightly decrease in the percentage, mainly because the difference of almost 40 percent of actions related to asset misappropriation recorded between 2011 and 2020 were divided to the other categories. From a qualitative point of view, it can be concluded that the main categories of actions related to the cyber-enabled fraud did not suffer major changes, the real difference being, probable, in the deeper substratum of each category, composed of the instruments, techniques and procedures used by the perpetrators.

Last but not least, the future, from the point of view of the impact of organized crime, in a broad sense, respectively of cyber fraud, in a narrower sense, seems to have strong shades of grey. This is mainly due to the innovations in the field of information and communication technology, respectively those in the fields related to cyberspace. The year 2019 was the cornerstone of two situations: the use of artificial intelligence to perform a complex action of cyber fraud, while reaching a new peak of technological development - the use of a new type of sophisticated computer to operate a calculation in minutes that, by using the most powerful computer in existence to date, it would have lasted about 10,000 years. It is important to analyze, in the future, the impact that these new technologies will have on the various fields of activity, mainly by identifying the threats that derive from the use of these innovations by actors with malicious and/or illicit personal intentions.

## **References**

- \*\*\* (2020) “Woman loses £320,000 in 'romance fraud' scam”. *BBC News*. 20 October. [online], <https://www.bbc.com/news/uk-england-somerset-54613937>, accessed on 25 November 2020.
- Alhassan, N.S., et al. (2018) “Salami Attacks and its Mitigation - An Overview”, *Proceedings of the 5<sup>th</sup> International Conference on “Computing for Sustainable Global Development”*, pp. 4639-4642.

- Button, M., Cross, C. (2017) "Technology and Fraud: The 'Fraudogenic' Consequences of the Internet Revolution". The Routledge Handbook of Technology, Crime and Justice. Publisher: Routledge.
- CERT-LT [online], <https://www.nksc.lt/en/reports.html>, accessed on 24 January 2021.
- CERT-PL [online], <https://www.cert.pl/en/publikacje/>, accessed on 24 January 2021.
- CERT-RO [online], <https://www.cert.ro/doc/ghid?page=2>, accessed on 24 January 2021.
- CSIRT-SK [online], <https://www.csirt.gov.sk/graf-2019-8af.html>, accessed on 24 January 2021.
- Danhieux, P. (2013) "Email Phishing Attacks", *The Monthly Security Awareness Newsletter for Computer Users*, February [online], [https://www.brockport.edu/support/information\\_security/cybersecurity\\_awareness/STC\\_cyber101\\_phishing.pdf](https://www.brockport.edu/support/information_security/cybersecurity_awareness/STC_cyber101_phishing.pdf), accessed on 25 November 2020.
- Detica (2011) "The cost of cybercrime", Guildford [online], [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/60943/the-cost-of-cyber-crime-full-report.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/60943/the-cost-of-cyber-crime-full-report.pdf), accessed on 03 December 2020.
- Durbin, S. (2020) "The Future's Biggest Cybercrime Threat May Already Be Here" [online], <https://www.darkreading.com/vulnerabilities---threats/the-futures-biggest-cybercrime-threat-may-already-be-here/a/d-id/1338439>, accessed on 04 December 2020.
- European Payments Council (2019) "2019 Payment Threats and Fraud Trend Report" [online], <https://www.europeanpaymentscouncil.eu/sites/default/files/kb/file/2019-12/EPC302-19%20v1.0%202019%20Payments%20Threats%20and%20Fraud%20Trends%20Report.pdf#page=16&zoom=100,72,426>, accessed on 28 January 2021.
- Gorham, M. (2019) "2019 Internet Crime Report" [online], [https://pdf.ic3.gov/2019\\_IC3Report.pdf](https://pdf.ic3.gov/2019_IC3Report.pdf), accessed on 02 December 2020.
- Jahankhani, H., Al-Nemrat, A., Hosseinian-Far, A. (2014) "Cybercrime classification and characteristics". *Cyber Crime and Cyber Terrorism Investigator's Handbook*, pp. 149-164.
- Line, M., et al. (2006) "Safety vs security?". *Proceedings of the 8th International Conference on Probabilistic Safety Assessment and Management*, New Orleans, USA.
- Moneva, A., et al. (2020) "Recorded Cybercrime and Fraud Trends in UK during COVID-19", *Statistical Bulletin on Crime and COVID-19*, Issue 6. Leeds: University of Leeds.
- Nurse, J., Bada, M. (2018) "The Group Element of Cybercrime: Types, Dynamics, and Criminal Operations". *The Oxford Handbook of Cyberpsychology*. Oxford: Oxford University Press.
- Nurse, J. (2018) "Cybercrime and You: How Criminals Attack and the Human Factors That They Seek to Exploit". *The Oxford Handbook of Cyberpsychology*. Oxford: Oxford University Press.
- Online: <https://dictionary.cambridge.org/dictionary/english/fraud>, accessed on 25 November 2020.
- PwC [online], <https://www.pwc.com/gx/en>, accessed on 25 January 2021.
- Reuters (2019) "Google claims its quantum computer solved a 10,000-year problem in seconds" [online], <https://www.cnn.com/2019/10/23/google-claims-successful-test-of-its-quantum-computer.html>, accessed on 04 December 2020.
- Sabillon, R. et al. (2016) "Cybercrime and cybercriminals: A Comprehensive Study", *International Journal of Computer Networks and Communications Security*, 4(6), pp. 165-176.

# AI-Powered Defend Forward Strategy

Jim Chen

U.S. Department of Defense National Defense University, Fort McNair, Washington, USA

[jim.chen@ndu.edu](mailto:jim.chen@ndu.edu)

DOI: 10.34190/EWS.21.505

**Abstract:** The goal of the defend forward strategy in the cyber domain is to thwart attacks at their sources or at least mitigate their impact before they reach their targets. To achieve this goal, specific capabilities must be built into the technology and the processes that support this mission. These capabilities include but are not limited to the following ones: robust intelligence collection, accurate decision-making, quick and accurate targeting, constantly changing to avoid being detected by adversaries, unexpected maneuvering to generate precise and surprising effect at the speed of light, and objective assessment of missions accomplished. It has to be acknowledged that the high demand for these unique capabilities cannot be satisfied without the employment of artificial intelligence (AI). This paper explores one way of building these unique capabilities utilizing AI in order to support the defend forward strategy. The proposed solution calls for the integrated architecture of capability, speed, and precision as well as the checks-and-balances architecture. The paper reveals how strategic advantages can be achieved via the use of these new capabilities supported by human-machine teaming. This exploration can provide guidance for developing new capabilities for commanders' toolkits. Consequently, it will help to nurture a cyber persistent force comprised of humans and machines.

**Keywords:** defend forward, persistent engagement, artificial intelligence, human-machine teaming, strategic advantages

---

## 1. Introduction

Challenges in cyberspace pose unique threats to national security. They are hard to be dealt with for the following reasons.

First of all, the entry level of cyber operations is low. The low-cost but effective cyber tools have assisted nation-states, criminal groups, terrorists, or disgruntle individuals in launching cyber operations to gain strategic advantages. With the help of these tools, adversaries gain access to victims' systems, disrupt normal processes, spread disinformation, steal information, destroy computing systems, networking systems, and infrastructure systems. As consequences of cyber operations lack human casualty and property damage in most cases, both decision-makers and operators of these operations usually escape severe punishment. Eventually, they may achieve strategic advantages with ways and means short of armed conflict. In General Nakasone's words, "[c]yberspace provides our adversaries with new ways to mount continuous, nonviolent operations that produce cumulative, strategic impacts ..... without reaching a threshold that triggers an armed response" (Nakasone 2019).

Besides, new technologies make it possible to dynamically manipulate the speed of cyber operations. To disrupt and/or destruct target systems, cyber operations can be conducted at the speed of light. Meanwhile, to collect intelligence, steal information, or support influence campaigns, cyber operations can be conducted in a hidden mode for a long period of time without being detected. Both types of cyber operations generate strategic surprise in their own ways.

In addition, a strategic impact may be generated from seemingly irrelevant, piecemeal, and nonviolent cyber events. It usually takes extra time and requires additional resources on the defense side to identify the relationships among these events and then to make sense out of them. This leaves an adversary with an upper hand as they can hide their real intention for some time and their operations are not immediately challenged.

Moreover, there are strategic decisions that must be made, sometimes within a very short period of time. The decision-makers or commanders have to answer the following questions: Is the cyber campaign to be launched in compliance with the international laws? Is the cyber campaign authorized? What are the 2<sup>nd</sup> and the 3<sup>rd</sup> order effects of the cyber campaign? Is it cost-effective in launching the cyber campaign? When is the most suitable time to start the cyber campaign? In which way should it be launched to achieve the effectiveness?

These are challenges that must be addressed. To deal with them, we have to examine the current cyber defense strategies and find out whether they work effectively or not. This assessment can help us to figure out ways of

overcoming the limitations and further improving the strategies. It can also motivate us to come up with new strategies.

First and foremost, the traditional cyber defense strategy that is only focused on defense proves to be not working as breaches have been reported even when the strategy is in place. As pointed out by Ravich and Cardon (2020), “[t]he growing use of cyber weapons against the United States, ranging from intellectual property theft, disinformation, data destruction, and denial of service attempts is a clear sign that a purely defensive strategy will fail.” As the traditional cyber defense strategy does not address the dynamic aspect of cyber operations, it is not capable of avoiding, detecting, mitigating, and blocking various types of cyber breaches or attacks.

To cope with these issues and to change the passive defense to an active one, the defend forward strategic concept was introduced in the 2018 United States (U.S.) Department of Defense (DoD) Cyber Strategy Summary, which states that DoD will “defend forward to disrupt or halt malicious cyber activity at its sources, including activity that falls below the level of armed conflict.” This strategy is echoed in the Cyberspace Solarium Commission Report published in March 2020. The report maintains that “the Commission integrates defend forward into a national strategy for securing cyberspace using all the instruments of power”. It states that to defend forward is to “proactively observe, pursue, and counter adversaries’ operations and impose costs short of armed conflict..... with all the tools at its disposal and consistent with international law”. Borghard (2020) holds that this strategic concept can “disrupt and defeat ongoing malicious adversary cyber campaigns, deter future campaigns, and reinforce favorable international norms of behavior”. Lloyd (2020) also claims that this strategy can “inflict[s] costs on bad actors”.

The defend forward strategy possesses advantages over the traditional defense strategy. However, it does have some issues that should be addressed. This paper intends to address them with the help of AI. It proposes a solution that calls for the integrated architecture of capability, speed, and precision as well as the checks-and-balances architecture. The paper reveals how strategic advantages can be achieved via the use of this new framework, which is also capable of addressing the limitations in AI.

This paper is structured as follows: Here in Section 1, the environment and the necessity of having cyber defend forward strategy are illustrated. In Section 2, the advantages and disadvantages of this strategy are discussed. In Section 3, requirements for implementing this strategy are examined. In Section 4, a trusted artificial intelligence framework is recommended to address the issues in the current cyber defend forward strategy. The benefits of this framework are discussed. Potential topics for future research are suggested. In Section 5, a conclusion is drawn.

## **2. The cyber defend forward strategy: Advantages and disadvantages**

In the 2018 DoD Cyber Strategy Summary, the goal of the defend forward strategy is clearly pointed out. It is to “disrupt or halt malicious cyber activity at its source, including activity that falls below the level of armed conflict.” To achieve this goal, activities in breadth and depth are needed. The breadth of activities involves operations outside the “blue space” (i.e. U.S. domestic cyberspace), in the “red space” (i.e. adversary cyberspace), and the “gray space” (i.e. everywhere else), as mentioned in the Council on Foreign Relations 100 Blog Post (Council of Foreign Relations, April 22, 2020). The depth of activities involves “operational preparation of the environment; prior intelligence collection and operations to identify vulnerabilities and exploits; the development or procurement of tools to deliver the intended effects; and the ability to hold targets at risk over time to deliver the appropriate effect on a decision-maker’s request”.

The defend forward strategy enjoys many advantages. Here, three of them are emphasized below:

- (1) Imposing cost

As claimed by Pomerleau (2019a), the defend forward strategy makes it possible to challenge “adversary activities wherever they operate” with the help of persistent engagement. The challenges include disrupting internet access, making the work of hacking harder at the very least, and other operations. These fast and surprising challenges may provide a certain level of deterrence, thus helping to shape the behavior of an adversary. Borghard (2020) also argues that this strategy is able to impose cost upon malicious actors rapidly at the desired time. The challenges imposed upon adversaries make it difficult for them to conduct malicious operations and campaigns, thus increasing their uncertainty about the likelihood of success. The challenges include but are not limited to countering an adversary’s “offensive cyber capabilities and infrastructure, the

organizations that support their cyber operations and campaigns, and the locus of their decision-making". In Sawmiller (2020)'s term, the cyber forces can identify the weaknesses of an adversary, learn the intentions and capabilities of the adversary, and "then take action to deny, degrade, disrupt, destroy, or manipulate (D4M) cyber infrastructure owned, operated, or controlled by the adversary".

- (2) Sharing information

Once in the gray space or the red space, information about an adversary and its capabilities can be collected and analyzed. Sharing this information with targeted owners and operators can enable their rapid cyber responses should there be a cyber attack. Also, sharing this information with relevant parties in public and private sectors as well as allies and partners may help them to get prepared before a cyber attack occurs. Once being prepared, the adversary's attempt will be blocked or at least its effect will be reduced.

- (3) Establishing norms

Smeets (2020) mentions that some researchers think that this strategy together with the persistent engagement strategy can help to establish norms. This is because an adversary now has to think twice before launching an attack since there are severe consequences. This eventually helps to promote stability and establish norms in cyberspace.

Nevertheless, there are some disadvantages that must be considered in order to implement this strategy. They are listed below:

- (1) Legal challenges

As Deeks (2020) points out, in principle, even if they are allowed, countermeasures are limited in scope in its usage. These limitations include but are not limited to non-use of force, the need for appropriate attribution, the general requirement to provide notice and negotiation prior to the use of countermeasures, the constraints of proportionality, and constraints of reversibility. These principles in international law apply to cyberspace. In Egan (2017)'s term, the purpose of countermeasures in response to an international wrong in cyberspace is not to punish but to compel the wrongdoing state to resume compliance with its international obligation. Schmitt and Vihul (2017) also state that countermeasures should not violate fundamental human rights or peremptory norms of international law. However, as argued by Deeks (2020), if cyber operations constitute a violation of international law such as the non-intervention rule, the victim state can take actions with the goal of persuading the state that launches the attack to stop these malicious operations. In such a case, the victim state has several options. It may choose not to inform prior to the use of countermeasures, to offer to negotiate, or to give "advance notes of its plan to impose countermeasures". However, "[c]ountermeasures must be nonforcible and proportional, and are limited to a temporary non-performance of the injured state's international obligations toward the wrongdoing state." One of the recommendations that Deeks (2020) provides is to specify which behaviors in cyberspace violate international law, make transparent the general policy about the use of countermeasures, and conduct proportional countermeasures with a multilateral approach. Beyond doubt, this requires efficient and effective risk management analysis that involves cost-benefit analysis. Ideally, the cost-benefits analysis helps to decide what measures should be taken in disrupting adversary capabilities and infrastructure, and in affecting the adversary's decision-making cycle. There is a cost associated with any measure taken. Whether it is worthwhile to utilize a specific measure should be determined by the overall gains both at the strategic level and at the operational level. In strategy making, this part should be considered. Here, the question is not whether a cost is needed or not. Instead, the question is how much cost should be imposed upon an adversary in a specific context to make the strategy effective. What is more, this kind of decision sometimes has to be made within a short period of time to stop the bleeding. These enhanced capabilities are not well addressed in the current version of the defend forward strategy.

- (2) Strategic challenges

The decision-makers have to be mindful of the ultimate goal of the defend forward strategy, i.e. to change adversary behavior. As well put by Borghard (2020), the defend forward strategy seeks to create costs for adversary on one hand and improve the situational awareness on the other hand. This is because "unlike in the realm of nuclear deterrence, in cyberspace we cannot expect a binary outcome – the use of a capability versus nonuse". Again, what should be figured out is which capabilities and how much capabilities should be used. The relatively precise calculation is essential for the success of the defend forward strategy. Unfortunately, this is also not well discussed in the current version of the defend forward strategy.

In addition, it needs to be pointed out that the implementation of this strategy may lead to unintended consequences in some cases. One of them, as argued by Smeets (2020), is the possible negative implication for the alliance. To improve the situational awareness, “working with partner nations to help secure their systems” is a good method, as mentioned by Pomerleau (2019b). Authorized by the National Defense Authorization Act (NDAA) 2019, Cyber Command defensive teams can “operate outside the DoD networks to help our allies defend forward”. Specifically, at the invitation of the host nations, these teams “work with them within their networks in a defensive role”. This makes it possible for these teams to get as close to adversaries as possible and to gain some tremendous insights into the strategies that adversaries use. According to Smeets (2020), if a cyber operation is launched in the networks of an ally or a partner, and if the U.S. cyber forces and the allied cyber forces have different priorities, say one party wants to keep gathering intelligence of adversarial activity while the other party attempts to disrupt and deny adversarial activity, “there could be a negative impact on intelligence operations and capabilities beyond these systems and networks” without appropriate coordination. Consequently, this may lead to “i) loss of trust due to offensive cyber effects operations in allied systems or networks; ii) compromise of allied intelligence operations and capabilities; iii) exploitability of the strategy by adversaries; and iv) the implementation (and justification) of persistent engagement by other countries”. This literarily shows the importance of coordination. Under some circumstances, coordination should be conducted within a very short period of time. It also requires compromise of one party or both parties for the achievement of a common goal. This requires a complete understanding of the environment from various perspectives and in multiple domains. This may indicate that in order for this strategy to be applied in an allied country, that country should also adopt the defend forward strategy. Needless to say, a coordinated operation within the shortest amount of time is critical in this situation to avoid friction and enhance trust.

- (3) Operational challenges

Speed and change in maneuver are always two critical factors. Once a decision is made, a quick response should be followed. Mears and Mariani (2020) closely examine the temporal dimension of defending forward. They hold that “relying on rapid response and reactivity may place challenging, if not impossible, demands on capability, thus there is a related need to invest in anticipation.” What is more, changes in tactics, techniques, and procedures (TTPs) should be introduced in every context if possible as the same tricks are not effective in most cases. Hence, new tricks have to be designed, developed, and executed based on the requirements of a specific context. Only the combination of these two critical factors can cause surprise effect, thus generating deterrence at the cyber level (Chen, 2018) and (Chen, 2017).

Besides, at the operational level, communications should be precisely handled. On one hand, secure and secret communications should be established and maintained. The defend forward strategy requires sensors and responding systems to be hidden in various networks across the world without being detected. However, necessary communications among sensors, responding systems, and command and control (C2) centers are still needed. How can this type of communications not be detected by adversaries in the gray space and in the red space? This is another challenge that has to be addressed. On the other hand, proper signals need to be sent to an adversary to avoid the escalation of a situation in some cases. At which point should such signals be generated? When should such signals be sent out? According to Healey (2019), only negative feedbacks lead to de-escalation of a situation. In the cyber domain, negative feedbacks may not be generated sometimes. As a result, “conflict could easily spiral into a war of attrition”, thus leading to the “new forever war in cyberspace”.

- (4) Resource challenges

The success of the defend forward strategy also depends upon resources, both personnel and material. How much resources are needed during normal times and how much resources are needed during a specific conflict should be figured out and allocated. Any change in situation can lead to the change in requirement for resources. Without sufficient resources, it is hard to guarantee a success in a maneuver or in a conflict. In this sense, figuring out sufficient resources needed in a prompt way makes a difference in competition or in conflict.

The defend forward strategy requires special capabilities, capacities, and authority. With special capabilities and authority, adversary cyber activity can be halted at its source or at least can be degraded before it reaches its target. With the special capacities and authority, the cyber forces can maneuver inside and outside their own networks. In Kollars and Schneider (2018)’s term, “the strategy places defense outside the bounds of the .mil (the networks owned and operated by the U.S. military) and instead advocates defense of resources that enable military operations but may operate on the .com (private industry). The strategy also expands departmental efforts beyond U.S. geographic boundaries.” Beyond doubt, extra resources are needed to support special

capabilities and capacities. Even with sufficient resources, it takes time to develop and prepare for special capabilities and capacities.

To summarize, cyber operations that disrupt, deny, or degrade adversary capabilities demand speed and precision. In the next section, requirements for implementing the defend forward strategy are examined in detail.

### 3. Requirements for implementation

The above study shows that in order to successfully implement the defend forward strategy, the following requirements must be satisfied. These include but are not limited to the following categories: capability, speed, and precision.

The capability category includes capability of collecting relevant intelligence, capability of processing big volume of data, capability of performing data analytics, capability of recognizing patterns, capability of identifying targets, capability of making decisions, capability of performing C2, and capability of sharing relevant information with other government agencies, the private sector, allies, partners, capability of maneuvering from the blue space to the gray space and/or the red space, capability of imposing cost, and capability of stopping attacks. It has to be mentioned that all these capabilities must be supported by speed and precision. Without these two critical dimensions, these capabilities are not full-fledged capabilities, and any operation that depends upon them is doom to fail. The relationship among these three categories is captured in Figure 1 below:



**Figure 1:** Integrated architecture of capability, speed, and precision

This figure indicates the dependency of a capability upon speed and precision in a specific environment. Hypothetically, in one case, one capability requires fast speed and high precision. In another case, another capability is fine with medium speed and high precision. No matter what the situation is, these three factors must be integrated in implementation. The architecture that integrates these three categories can help to set up a priority list when multiple requests are evaluated at the same time.

This integrated architecture can also be used in analyzing the challenges discussed in the previous section. With respect to a legal requirement, it is fine to have slow speed and highest precision when establishing a legal principle or rule, but it requires fast speed and high precision for the enforcement of an adopted legal principle or rule. With respect to a strategic requirement, imposing cost on an adversary may require fast speed and high precision in one case but may require medium speed and high precision in another case. Meanwhile, coordination and trust building usually take a long time and may cover a broad area. Hence, it may go with medium or low speed and medium or low precision. With respect to an operational requirement, fast speed and high precision are always required no matter whether it is for maneuver, change, or communications. With respect to a resource requirement, fast speed and high precision are needed for the delivery of resources, either personnel or material. However, it has to be acknowledged that the training or education of personnel may take low to medium speed.

One effective means of ensuring fast speed and high precision is AI. In comparison, human manual processes can seldom reach that level of speed and precision in accomplishing tasks. Hence, AI systems and applications should be called in.

To support the defend forward strategy, an AI system, in any given situation, should be able to figure out the most efficient, effective, and proportional course of action, be able to maneuver below the threshold of armed conflict while imposing cost onto an adversary, be able to make changes constantly both in appearance and in payloads, and be able to earn trust of allies or partners. All these require the cost-benefit analysis. The calculation in this analysis needs AI to support speed, precision, and completeness.



#### **4. The trusted artificial intelligence framework**

In contexts like this, it is hard to imagine how these requirements can be satisfied without the involvement of AI. As pointed out by Chen (2019), AI possesses superiority of speed and precision. It is capable of speedy calculation, fast data analysis, swift pattern recognition, rapid processing, and quick learning. Besides, it is tireless, with no emotions, feelings, wants, and needs. It makes automation and autonomy possible in many fields where humans are involved, such as decision-making and execution of action. Beyond doubt, AI can address the challenges with respect to capability, speed, and precision. It is able to collect relevant intelligence, process big volume of data, perform data analytics, recognize patterns, identify targets, make decisions, perform C2, share relevant information with other parties, maneuver between different spaces, impose cost, and eventually stop attacks.

What should be considered next is how AI can be used ethically and responsibly as research shows that it has some intrinsic limitations. Chen (2019) holds that its limitations include but are not limited to the lack of sufficient knowledge of law and ethics, lack of accountability, and potential bias throughout the whole process. Zais (2020) also states, “[s]ome defense scholars have advocated a smarter military, emphasizing intellectual human capital and arguing that cognitive ability will determine success in strategy development, statesmanship, and decisionmaking. AI might complement that ability but cannot be a substitute for it.” What also should be considered is how adversarial AI should be dealt with so that AI systems do not fall as victims of malicious attacks.

An ethical and responsible AI framework is a solution to these challenges because it not only possesses the speedy and precise capabilities enabled by the integrated architecture of capability, speed, and precision but also restricts the use of these capabilities for legitimate purpose in a responsible way with the help of the checks-and-balances architecture.

First of all, various capabilities should be designed and developed. These capabilities include but are not limited to the capabilities of collecting relevant intelligence, processing big volume of data, performing data analytics, recognizing patterns, identifying targets, making decisions, performing C2, sharing relevant information with other parties, maneuvering between different spaces, imposing cost, and eventually defeating attacks; the capabilities of constantly and dynamically changing appearance and payloads while maneuvering between different spaces and/or hiding in one specific space without being detected by adversary cyber forces; and the capabilities of secretly communicating with a C2 center in a direct or an indirect way.

To support fast speed in processing, both supervised learning mechanism and the back-end processing are utilized. The relevant international law and ethics are translated into legal principles and rules as well as ethical principles and rules, which serve as guidance for all cyber operations. Specifically, these principles and rules are coded, and pre-loaded into reference libraries accessible by all programs/codes in AI systems. From these principles, new rules can be derived given specific triggers in input datasets, thus enriching the reference libraries. The pre-loaded reference libraries certainly speed up the processes. Likewise, while a dataset is processed at the front end, its metadata and its legitimacy are checked and processed simultaneously at the back end. Should there be a violation detected at any point, the whole processing is halted and the dataset is dropped into the location where untrusted datasets are stored. Should a dataset be legitimate and validated, various potential courses of action, their significance weights, their risk likelihood, their risk impacts, their success rates, and their pros and cons are all calculated at the back end and then loaded into an AI system. This certainly speeds up processing and analysis. Besides, unsupervised learning mechanism is used to find new patterns from input datasets. This helps to figure out new tricks, which can be coded and added into reference libraries.

To support precision in processing, a target is never randomly selected. Instead, it is identified with sufficient data points. To address the bias issue, a target selected is always verified and confirmed. The verification and confirmation are conducted by other human-machine teams. This guarantees precision in most cases. As for a selected course of action, it should pass multiple rigid tests and survive the cost-benefit analysis as well as the verification and confirmation conducted by human-machine teams.

Next, to ensure ethical and responsible use of these capabilities, the checks-and-balances architecture that Chen (2019) proposes is employed. Empowered by human-machine teaming, this architecture explicitly differentiates the roles and responsibilities of humans and machines, thus clearly defining the accountability for any action. In

this architecture, dynamic human supervision and involvement are built-in features. Besides, both the front-end process and the back-end process are utilized to speed up processing. Ultimately, it ensures fast speed, completeness, precision, reliability, and dynamics simultaneously. The collaboration among humans and AI systems helps to build trust, which is further enhanced through positive feedback loops assisted by the reinforcement learning mechanism. In this architecture, there are the law/ethics/rule-making component, the judicial component, and the execution component. Each component consists of a human-machine team. Each decision made in one component is verified and confirmed by the human-machine teams in other components. If there is a conflicting view, experts from different fields are called in to the judicial component to make a judgment. A final decision is made based on the judgment. It is in this way that ethical and responsible use of these capabilities are guaranteed.

A significant effect can be generated when all these components are precisely put together into a trusted AI framework. Below, we are going to examine how this framework can address the challenges discussed in Section 2.

This framework proposed here is capable of handling legal challenges. An AI system can quickly select or create relevant courses of action as countermeasures that are not only proportional but also short of armed conflict. Among these courses of action, one will be selected by the human and machine team of the executive component based on a specific context. This selection is further verified and confirmed by the human and machine teams of both the law/ethics/rule-making component and the judicial component prior to its execution. This guarantees trust and precision. As the knowledge of law and ethics has been coded as legal and ethical principles and rules within reference libraries accessible by all programs or codes in AI systems, it can be quickly checked and compared. In addition, with the knowledge of law and ethics as well as various potential courses of action pre-loaded into AI systems, a selection can be quickly made should the same conditions be met in a case under investigation. This also guarantees speed. Collectively, the legal challenges can be successfully dealt with.

This framework is capable of tackling strategic challenges. It is good at calculating the cost imposed onto an adversary for each capability, so it can figure out which capabilities should be used, how much capabilities should be employed, when and how they should be executed. All these are calculated in the cost-benefit analysis. The results of the analysis are prioritized courses of action with associated costs. Besides, the risk likelihood percentage, the risk impact value, the corresponding 2<sup>nd</sup> and 3<sup>rd</sup> order effects of each option are provided to make the decision-making process interpretable. The final selection is verified and confirmed by the human-machine teams in other components. These courses of action can be either virtual or physical, or even both, determined by the requirements in specific environments. As they are well calculated and proportional, they may not cause an escalation of the situation. However, they may carry and convey strong warning messages. Since a trusted AI system can calculate the gains and losses of each potential maneuver, it should avoid any potential conflict of interest and select a coordinated maneuver option. This, of course, will enhance trust since allied intelligence operations and capabilities are not compromised.

This framework is capable of dealing with operational challenges. It is able to quickly create polymorphic and even metamorphic changes to TTPs. The constant change of appearance and payloads may deceive adversaries and cause the failure of their detection systems. Instead of utilizing one tool in many places, new tricks can be used in every place, at least it seems to be. Ultimately, surprise effect and deterrence can be generated at the cyber level. This will certainly put our cyber forces in an advantageous position. Together with the use of hidden channels, these capabilities can also support secure and secret communications. The polymorphic and metamorphic capabilities can make sensors and responding systems hard to be detected by adversaries. They can also make changes to hidden channels constantly. Besides, the hidden sensors can help generate negative feedbacks should it be needed, thus preventing the escalation of a situation.

This framework is capable of managing resource challenges. It can quickly figure out how much personnel and material resources are needed for one specific maneuver, how the resources should be allocated and delivered in the most efficient way, and when the resources should be delivered. Likewise, after multiple courses of action are generated, one of them is selected. This selection is verified and confirmed by the human-machine teams in other components to assure precision.

As shown above, the trusted AI framework can successfully address the challenges of the defend forward strategy and further empower the strategy. With the support of trust, speed, precision, unique capabilities can make a big difference. Consequently, the purpose of the defend forward strategy can be achieved: Adversary cyber activity can be halted at its source or at least can be degraded before it reaches its target.

Clearly, the trusted artificial intelligence framework is advantageous. Future research should figure out an optimal way of implementing this framework by building a prototype. Thus, the existing capabilities can be tested in various environments and fine-tuned; new capabilities can be created; and the weaknesses of the framework can be identified and overcome. Only by so doing can this framework be further improved.

## 5. Conclusion

The purpose of the defend forward strategy in the cyber domain is to thwart attacks at their sources or at least mitigate their impact before they reach their targets. It has its advantages but it also has its challenges. In this paper, the challenges within the legal, strategic, operational, and resource aspects are closely examined.

To address these challenges, the trusted artificial intelligence framework is recommended in this paper. This framework consists of the integrated architecture of capability, speed, and precision as well as the checks-and-balances architecture. It is also supplemented by the cost-benefit analysis. With such a unique design, this framework can make the defend forward strategy do the magic with trust, speed, and precision.

This framework can seamlessly support robust intelligence collection, accurate decision-making, quick and accurate targeting, constantly changing to avoid being detected by adversaries, unexpected maneuvering to generate precise and surprising effect at the speed of light, and objective assessment of missions accomplished.

This paper explores one way of enhancing the defend forward strategy. It reveals how strategic advantages can be achieved via the use of new capabilities that take advantage of the best that both humans and machines can offer. This exploration can provide guidance for developing new capabilities for commanders' toolkits. Consequently, it will help to nurture a cyber persistent force that is comprised of humans and machines.

## References

- Borghard, E. (2020) "Operationalizing defend forward: How the concept works to change adversary behavior", *Lawfare* (March 12, 2020). Retrieved from <https://www.lawfareblog.com/operationalizing-defend-forward-how-concept-works-change-adversary-behavior>.
- Chen, J. (2019) "Who Should Be the Boss? A Machine or a Human?", *Proceedings of the European Conference on the Impact of Artificial Intelligence and Robotics*, pp.71-79.
- Chen, J. (2018) "On Levels of Deterrence in the Cyber Domain", *Journal of Information Warfare*, Vol.17, No.2, pp.32-41.
- Chen, J. (2017) "Cyber Deterrence by Engagement and Surprise", *PRISM*, Vol.7, No.2, pp.101-107.
- Council on Foreign Relations 100 Blog Post (2020) "U.S. Cyber Command's Malware Inoculation: Linking Offense and Defense in Cyberspace" (April 22, 2020). Retrieved from <https://www.cfr.org/blog/us-cyber-commands-malware-inoculation-linking-offense-and-defense-cyberspace>.
- Cyberspace Solarium Commission. (2020) "Cyberspace Solarium Commission Report". Retrieved from <https://www.solarium.gov/report>.
- Deeks, A. (2020) "Defend Forward and Cyber Countermeasures", *Public Law and Legal Theory Paper Series 2020-59*, University of Virginia School of Law.
- Egan, B. (2017) "International Law and Stability in Cyberspace", *Berkeley Journal of International Law*, Vol 35, 169, 178.
- Healey, J. (2019) "The Implications of Persistent (and Permanent) Engagement in Cyberspace", *Journal of Cybersecurity*, 2019, pp.1-15. DOI: 10.1093/cybsec/tyz008.
- Kollars, N. and Schneider, J. (2018) "Defending Forward: The 2018 Cyber Strategy Is Here", *War on the Rocks* (September 20, 2018). Retrieved from <https://warontherocks.com/2018/09/defending-forward-the-2018-cyber-strategy-is-here/>.
- Lloyd, W. (2020) 'Supporting the DoD's Defend Forward Initiative', RedSeal. Retrieved from <https://www.redseal.net/supporting-the-dods-defend-forward-initiative/>.
- Mears, A. and Mariani J. (2020) "The Temporal Dimension of Defending Forward: A UK Perspective on How to Organize and Innovate to Achieve US Cyber Command's New Vision", *The Cyber Defense Review*, The Army Cyber Institute, Vol.5, No.1, pp.55-74.
- Nakasone, P. (2019) "A Cyber Force for Persistent Operations", *Joint Force Quarterly*, Vol.92, 1<sup>st</sup> Quarter 2019, pp.10-14.
- Pomerleau, M. (2019a) "Two years in, how has a new strategy changed cyber operations?". Retrieved from <https://www.fifthdomain.com/dod/2019/11/11/two-years-in-how-has-a-new-strategy-changed-cyber-operations/>.
- Pomerleau, M. (2019b) "Here's how Cyber Command is using 'defend forward'". Retrieved from <https://www.fifthdomain.com/smr/cybercon/2019/11/12/heres-how-cyber-command-is-using-defend-forward/>.

**Jim Chen**

- Ravich, S. and Cardon, E. (2020) "Defending Forward in the Cyber Domain", *Defending Forward: Securing America by Projecting Military Power Abroad*, B. Bowman (ed.), Washington DC: FDD Press, pp.90-92.
- Sawmiller, J. (2020) "Fighting Election Hackers and Trolls on Their Own Turf: Defending Forward in Cyberspace". Retrieved from <https://digitalcommons.law.uidaho.edu/idaho-law-review/vol56/iss2/13/>.
- Schmitt, M. and Vihul, L. (eds.) (2017) *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2<sup>nd</sup> Edition), Cambridge, UK: Cambridge University Press.
- Smeets, M. (2020) "U.S. Cyber Strategy of Persistent Engagement & Defend Forward: Implications for the Alliance and Intelligence Collection", *Intelligence and National Security*, Vol.35, No.3, pp.444-453. Retrieved from <https://doi.org/10.1080/02684527.2020.1729316>.
- The U.S. Department of Defense. (2018) "Summary: Department of Defense Cyber Strategy 2018", Washington DC. Retrieve from [https://media.defense.gov/2018/Sep/18/2002041658/-1/1/CYBER\\_STRATEGY\\_SUMMARY\\_FINAL.PDF](https://media.defense.gov/2018/Sep/18/2002041658/-1/1/CYBER_STRATEGY_SUMMARY_FINAL.PDF).
- Zais, M. (2020) "Artificial Intelligence: A Decisionmaking Technology", *Joint Force Quarterly*, Issue 99, 4<sup>th</sup> Quarter, pp.71-73.

# Global Military Machine Learning Technology Development Tracking and Evaluation

Long Chen<sup>1,2</sup> and Jianguo Chen<sup>31</sup>

<sup>1</sup>Beijing Key Laboratory of Network Technology, Beihang University, Beijing, China

<sup>2</sup>Innovation Technology Research Institute, Beijing Topsec Network Security Technology Co Ltd, China

<sup>3</sup>Hebei Seismological Station, Earthquake Administration of Hebei Province, Shijiazhuang, China

[zhuanjiatuijian@126.com](mailto:zhuanjiatuijian@126.com)

DOI: 10.34190/EWS.21.092

**Abstract:** We have carried out global military machine learning technology development tracking and evaluation research, summarized the global military machine learning technology development status, analyzed its main technology development path, studied its development trend, analyzed the global military machine learning technology typical military application cases and development prospects, and proposed Enlightenment suggestions. Related institutions are paying close attention to the development strategy of machine learning in the military field. Representative countries have formulated the technical route of machine learning in the military field. In particular, the U.S. military has seized the opportunity for Intelligence construction. During the past few years, it has been conducting theoretical preparations and technological evolution. The U.S. military's intelligent construction is speeding up in an all-round way, and the overall combat capability will make a sharp jump. In particular, the U.S. Department of Defense(DoD) has accelerated the militarization of artificial intelligence applications and specially established the Joint Artificial Intelligence Center (JAIC) to coordinate and advance military research on artificial intelligence. With regard to machine learning to strengthen the construction and operations of various services and arms, militaries have intensively deployed various military intelligence research projects, carried out research on machine learning intelligent algorithms and promoted the transformation of artificial intelligence technology to intelligence processing, unmanned platforms, command and control, and weapon equipment systems. Troops from different countries around the world are taking machine learning technology into their land-based, sea-based, air-based, space-based and network space platforms weapons, networks and other systems. Taking the US military as an example around machine learning, the US Army conducts research on distributed processing and applied machine learning systems in autonomous networks and heterogeneous environments. The Navy develops unmanned naval information and response electronic attack projects. The Air Force's "quantum plan" and autonomous clusters resilient network, machine learning wingman, and six-generation machine developed so that it greatly increased combat power. Marines carry out the depth of reinforcement learning collaborative information warfare. Space army carries out analysis of the space-based data management. In particular, a series of planned network covered troops cyber threat defense, military IoT network defense, machine learning behavior detection, social network data analysis, and network electronic warfare, and other dimensions. In addition, we investigated the induction machine learning in future operations, intelligence, network, logistics, identification, health, trend data, and a plurality of key areas of current situation with development trend. We also put forward the suggestions.

**Keywords:** global military, machine learning, artificial intelligence, project tracking

---

## 1. Introduction

In the new round of research and application of machine learning craze, world military powers-related institutions have to track and evaluate the technical high ground machine learning and strengthen strategic consulting, technology research, and military applications efforts. Each country's national defense departments develop machine learning research applications at the same time. Machine learning also enhance the effectiveness of the various branches construction and operational aspects gradually multiplier from operations, networks, reconnaissance, logistics, medical, simulation, and data on the situation

The future development of military technology has extremely high value and research significance. Because of this, we carried out global military machine learning technology development tracking and evaluation research, summarized the global military machine learning technology development status, analyzed its main technology development path, studied its development trend, analyzed the global military machine learning technology typical military application cases and development prospects.

---

<sup>1</sup> Corresponding author: [zhuanjiatuijian@126.com](mailto:zhuanjiatuijian@126.com) (Jianguo Chen)

## **2. Military background analysis**

In recent years, artificial intelligence is increasingly widely recognized to be able to "change the future war" disruptive technology direction. The world's military powers compete fiercely around the commanding heights of artificial intelligence technology. The United States has successively introduced artificial intelligence national strategies and military strategies, established core units for the implementation of artificial intelligence military strategies, and actively developed intelligent military equipment to capture a new round of military power absolute advantage over the competition.

At the same time, Russia, Japan, India and other countries also have developed artificial intelligence to enhance the development of national and military strategic level, innovation and infrastructure investment to increase artificial intelligence, to carry out theoretical and applied research in the field of artificial intelligence and enhance military combat capability based on artificial intelligence.

Throughout the development of the US military in various fields, it basically has followed a relatively scientific development method, including the formulation of a strategic system, the building of core forces, the support of capabilities and means and the implementation of combat missions. At present, the US military is gradually improving its artificial intelligence military strategy, building an artificial intelligence military talent training system, promoting the development of artificial intelligence military-civilian integration and actively developing artificial intelligence-based means in various military fields.

## **3. Artificial intelligence strategy evaluation**

According to the US National Defense Strategy (NDS), the US Department of Defense (DoD) established the Joint Artificial Intelligence Center (JAIC) and released Artificial Intelligence Strategy for US DoD (Mattis, 2018a). The U.S. DoD's artificial intelligence strategy mainly describes four strategic critical areas of planning.

### **3.1 Increase the establishment of artificial intelligence research projects for military scenarios**

In the entire US defense body system, they increase military research projects in software development and Internet research technology demonstration efforts, reserve strategic data resources in the field of artificial intelligence, create and provide a safe controlled massive defense data. They formulate AI policies standards and use AI to optimize the daily business workflow of the US military.

### **3.2 Cooperate with leading US civilian technology sector and academia in allied countries**

With the introduction of leading business of artificial intelligence, they ensure the safety situation of their country top inside and outside of the top technology manufacturers cooperation. They make full use of artificial intelligence algorithms form a military solution and through the development is carried out in co-operation with leading businesses, manufacturers and academic institutions in the fields of artificial intelligence to ensure national security.

### **3.3 Vigorously train military artificial intelligence technicians**

Although the Artificial Intelligence can not replace people like, it can reduce repetitive tasks. By artificial intelligence technology training US forces, they strengthen computer language learning technology in the defense system. And the AI training is popular with staff and improves recruitment.

### **3.4 Research on military ethics and artificial intelligence security**

AI's ethical, moral and legal are related to the United States national military development direction. US Joint Center for Artificial Intelligence will ensure full discussion to resolve ethical issues with the use of AI between military and non-traditional partners.

## **4. Current status of military research on machine learning**

The military application of artificial intelligence (AI) will profoundly affect the outcome of future wars, and it is also related to a country's international status. Relevant institutions of the world's military powers have tracked and evaluated the commanding heights of machine learning technology, and continuously strengthened consulting, research and application efforts. Tracking and evaluating the global military machine learning strategy shows

that relevant agencies, especially in the United States. And we pay close attention to the development of evaluation machine learning in the military field (Table 1).

**Table 1:** Military machine learning consulting evaluation

Mechanism	Years	Jobs	Main idea	Goals
Land (NDRI)	2019	Department of Defense AI Situation Assessment (Tarraf etc, 2019)	Evaluation and research on AI technology research and development, growth and investment in combat applications	Strengthen and improve the AI posture of the Department of Defense
U.S. Department of Defense (DoD)	2018	The technical roadmap of machine learning for the US military in future cyber operations (Mattis, 2018a)	The US military needs to rely on artificial intelligence as a defense tool. The United States will increase its attention to the defensive cybersecurity of hardware and software platforms as a prerequisite for the safe use of artificial intelligence. Ensure the safety, reliability and robustness of the AI system of the Ministry of Defense.	More than half of the new challenges and plans announced by relevant US research departments involve machine learning or predictive analysis.
U.S. Congressional Research Service (CRS)	2018	Fusion between robots, autonomous systems, and artificial intelligence (Hoadley, 2018)	The US military is worried that its opponents will use the Lethal Autonomous Weapon System (LAWS) against the US military. Assess the potential impact of autonomous systems and artificial intelligence on U.S. ground forces	Develop and implement autonomous systems and artificial intelligence strategies, clarify short-term, medium-term and long-term priorities, and cooperate in multiple areas
U.S. United Artificial Intelligence Center (JAIC)	2018	Coordinating machine learning consulting work (Bastian, 2020).	The U.S. military is advancing the core focus of its machine learning strategy. It aims to promote machine learning to quickly enable key operational tasks, establish a general machine learning infrastructure for the Department of Defense, coordinate machine learning activities at the Department of Defense level, and attract and cultivate first-class Talent team	Coordinate the development of the U.S. military's unmanned equipment and assist it to continue to maintain asymmetric advantages in new technical fields <sup>[5]</sup>
US Defense Innovation Unit (DIU)	2018	Machine learning strategy consulting to help the U.S. military quickly utilize the latest commercial technology (Mori, 2018)	Focus on military technological innovation within the Ministry of National Defense, while DIU pays more attention to relying on external innovation forces, especially commercial technological innovation to promote defense innovation and integrate	The civil sector commercial machine learning and data science experience and depth combined military operations, priority to promoting the work in computer vision, large database analysis and forecasting and strategic reasoning in three areas <sup>[7]</sup>

**Long Chen and Jianguo Chen**

Mechanism	Years	Jobs	Main idea	Goals
			into the innovation ecosystem	
Russian Government Military Industry Committee	2018	Equipment plan before 2025	The Russian Ministry of Defense established a special committee for the development of military and special robotic technology and equipment. The "Automation of the Armed Forces of the Russian Federation" military technology conference is held every year. According to Defense Minister Sergei · Shaoyin ancient name, the last three years the Russian armed forces set up 10 large research institutes and centers. These research institutes and centers are conducting research in various fields including artificial intelligence, robotics and drones.	By 2025, 30% of military technology and equipment will be automated. Before 2030 obtained from the remote control and the robot 30 percent of the combat forces.
Ministry of Digital Development of Russia	2020	Preliminary list of AI projects (Эльяс, 2020).	The government will allocate funds to government agencies implementing AI projects as part of its digital transformation plan. The total cost of the agency's digital transformation plan may be as high as tens of billions of rubles	The list will be implemented in four ministries and three government departments in Russia. The list details the projects to be implemented from 2023 to 2024 .
Meeting of the Council of the Russian Ministry of Defense	2020	Future development plan for artificial intelligence control weapons (TASS, 2020)	The future use of artificial intelligence in weapon control will largely determine the outcome of war.	Russia's Defense Ministry identified the five main priorities in the near future, which The fifth focuses on the development of weapons and equipment with artificial intelligence elements, including robotic systems, unmanned aerial vehicles and automatic control systems .

The RAND Corporation of the United States conducts a situation assessment of the DoD's AI. Through independent assessments, it points out wrong perceptions and makes policy recommendations to strengthen and improve the DoD's AI situation (Tarraf et al. 2019).

US DoD released a future network warfare of machine learning technology roadmap, increasing the defensive network security hardware and software platforms upgrades, to ensure that the DoD AI system is safe, reliable and robust (Mattis 2018b).

The U.S. Congressional Research Service studies the integration of robots, autonomous systems, and artificial intelligence. It outlines some of the potential impacts and military applications of autonomous systems and artificial intelligence in the U.S. ground forces, and points out a series of issues that the U.S. Congress needs to consider (Hoadley and Lucas, 2018).



The United States JAIC coordinated machine learning research work, promoted machine learning to quickly empower key operational tasks, established a general machine learning infrastructure for the DoD, coordinated machine learning research at the DoD level, and attracted and cultivated first-class talent team (Hoadley and Lucas, 2018).

US Defense Innovation Unit integrated into the Silicon Valley innovation ecosystem's priority is to promote the work in computer vision, large database analysis and forecasting, and strategic reasoning in three areas by introducing the latest machine learning techniques to help American business fast development of machine learning equipment (Mori, 2018).

Military Industrial Commission for the Russian government issued its AI equipment plan 2025 years ago. During the last three years, the Russian armed forces set up ten large research institutes and centers. They researched various fields including artificial intelligence, robotics, and drones, to conduct research on military technology and equipment automation.

2020 December 16th, the Russian Ministry for Digital Development released the AI preliminary list of items which will be implemented in four ministries and three government departments in Russia. The list details the projects to be implemented from 2023 to 2024. The government will allocate funds to government agencies implementing AI projects as part of its digital transformation plan. The total cost of the agency's digital transformation plan may be as high as tens of billions of rubles (Эльяс, 2020).

At the meeting of the Council of the Russian Ministry of Defense on December 21, 2020, Russian President Putin stated that he was convinced that the future use of artificial intelligence in weapons control will largely determine the outcome of the war. The meeting also identified the Russian Defense Ministry's five main priorities for the near future, which the fifth focuses on the development of weapons and equipment with artificial intelligence elements, including robotic systems, unmanned aerial vehicles and automatic control systems (TASS 2020).

In short, a series of measures have been taken to promote the military application of artificial intelligence, such as focusing on the top-level design of the national strategy and issuing a series of intelligence development strategic plans. They deploy various military intelligence research projects. They focus on breakthroughs in core and key technologies to lay a solid foundation for military intelligence development. In particular, the U.S. DoD has accelerated the militarization of artificial intelligence applications, and actively improved the functional efficiency of the U.S. DoD through its potential through empowerment rather than replace military personnel through artificial intelligence information systems. And they especially established the JAIC to coordinate and promote the military research of artificial intelligence.

In the military application of machine learning at this stage, especially the extensive and in-depth research on unmanned systems, machine learning has shown strong superiority. Simultaneously, in the future, human machine fusion intelligence will have more significant advantages than machine learning. Human-machine fusion intelligence will occupy an increasing proportion in the military field. Human-machine fusion intelligence will also urge the rapid development of new technologies in the military field to achieve high levels of humanity. The integration of aircraft will also be an inevitable requirement for the development of the military field.

Especially the human mind and cognitive style as intersubjectivity made the AI carry out their duties and promote each other. The intelligent man-machine fusion has in the military field broad prospects.

At present, most countries are at the stage of technical and theoretical preparations, and the US military has seized the lead in AI construction. Various signs indicate that the overall informatization construction of the U.S. military is about to be completed. Artificial intelligence technology has entered a mature period and is widely used in the civilian field. The U.S. military's intelligence construction has begun to accelerate. The goal is for most countries to follow the US Military's lead in the use of Artificial Intelligence by 2045. Once breakthroughs occur in key technologies, the U.S. military's AI construction will be accelerated in an all-round way, and the overall combat capability will be greatly improved. Global military machine learning research prospects and cases are shown in Table 2.

**Table 2:** Global military machine learning research application cases and prospects

Country	Mechanism	Project
United States	U.S. Department of Defense	ALPHA Intelligent Beyond Visual Range Air Combat System "Project Maven" drone computing vision Future ground forces man-machine formation
	U.S. machine learning algorithm war cross-functional team	Tactical unmanned aerial vehicles and air in full motion video (FMV) Research on computer vision algorithms for target detection, classification and early warning Convert massive data into actionable intelligence
	US DARPA application of next-generation machine learning-based	Adaptive behavioral learning electronic warfare (BLADE) Adaptive Radar Countermeasures (ARC) Explainable machine learning Modernization of high-performance computing Small sample adversarial machine learning Spectrum coordination challenge
Russia	Russia established ten large-scale artificial intelligence research institutes and centers	Humanoid robot Robot Force Long-range strike drone Artificial intelligence and drone swarms Unmanned driving system
United Kingdom	British Ministry of Defence (MoD)	Artificial intelligence technology tracking global radar The man-machine interface enables the military to improve the accuracy and efficiency Search for mines through automatic submersibles
India	Related Indian Army	Strengthen the construction of the military's network-centric warfare capabilities "Tactical command, control, communication and information system" construction plan Network-centric warfare begins to move towards intelligence
Japan	Japan Defense Equipment Agency	"Unmanned" equipment in three areas: land, surface, underwater, and air Actuator technology accelerates unmanned control of equipment Artificial intelligence builds analysis tools for ship automatic identification devices Automation and efficiency of UAV surveillance and reconnaissance".
Israel	Israeli military deployment department	Fully automatic robots used in military deployment Self-driving cars for border patrol Mixed formation of robots and soldiers Smart goggles receive medical guidance remotely and provide emergency assistance Deployed cameras Sensors installed on the tank Data collection Information can be shared data for processing Artificial intelligence and scene matching technology

## 5. Machine learning project tracking in various military fields

Machine learning is also gradually forming the effect on strengthening the construction of various services and arms (Jian and Xianghua 2020). At present, some countries use intelligent military robots for combat experiments. The military application of artificial intelligence has attracted widespread attention. When we look at how to use machines correctly, it is worth thinking about learning techniques to carry out correct and controllable military operations without causing negative effects.

Compared with other countries, the US military's machine learning projects in the military field have a leading advantage in terms of the quantity of equipment research and deployment and the depth of new technology research and development. Here we focus on tracking and researching current US military machine learning projects.

It can be seen that the US military has intensively deployed various military intelligence research projects, carried out machine learning intelligent algorithm research, and promoted the penetration and transformation of

artificial intelligence technology into intelligence processing, unmanned combat platforms, command and control, weapon equipment systems, and changes in combat methods (Table 3).

**Table 3:** Service cases of military machine learning applications

Service arms	Project	Content
Army	Information collection for autonomous network projects	Automatic network decision security Small sample learning Adaptive and autonomous defense against cyber attacks
	Distributed processing plan in heterogeneous tactical environment (DPHTE)	Fog computing platform Integrate smart edge device data
	Applied machine learning system	Reduce cognitive burden Small sample target detection Prioritize data Automatically filter out unimportant signals
	Russian Ministry of Defense tests artificial intelligence targets for military training	Modern aiming complex Autonomous targets with artificial intelligence algorithms Tracking system can simulate allied or enemy forces
Navy	Global Unmanned Naval Information Processing Project (TOPGUN)	Detect farther targets Identify more types of ships Water display performance
	Responsive Electronic Attack Measures Project (REAM)	Radar uses machine learning to perceive the environment Change its transmission characteristics Pulse processing algorithm to defend against electronic interference
	Russia's sea surface automatic analysis system	Satellite and terrestrial information systems Radar station data and internal channels to transmit information Receive data from over-the-horizon radar
Air force	Air Force "Quantum Project"	Computational Intelligence Autonomous reasoning and decision making Relevant social impact
	Autonomous cluster elastic network project (DPAA-SEA)	Research on autonomous cluster elastic network (DPAA-SEA) The distributed phased array antenna system of the project
	Machine Learning Wingman Project (Skyborg)	Independent, can attack and open the UAV
	Sixth generation machine	Computer miniaturization Machine learning technology Collection, compilation and processing ISR data Real-time analysis Deep reinforcement learning
	TACE system	Deep reinforcement learning algorithm Control the aircraft No human intervention required
	U.S. Air Force Industry-University Integration Innovation Agency AFWERX	Artificial intelligence virtual base Create comprehensive global coverage of intelligence, surveillance and reconnaissance (ISR) system system Identify cutting-edge commercial satellite payload concepts, designs and prototypes
	Russia's large attack drone	Technical vision, autonomous driving capabilities, and potential manned and unmanned collaboration with soldiers on the battlefield Volume_up

Service arms	Project	Content
		Content_copy Share Star_border
Marines	Deep reinforcement learning technology is integrated into USMC 's weapon system	Deep reinforcement learning technology Integrated into USMC Weapon system
Space Army	On-board data processing and interpretation	Space -based surveillance system All day Continuous work
	Real-time high-efficiency analysis of space remote sensing images	Cloud system and software for supercomputers Machine learning and artificial intelligence algorithms Raw data from space observation sensors Analysis of current affairs images
Cyber force	Large-scale Web Search Project (CHASE)	Develop, demonstrate and evaluate new automated cyber defense tools Machine learning and cyber attack modeling Automatic detection and defense against currently undetected advanced threats
	Military IoT cyber defense platform DeepArmor	Machine learning to identify and analyze unknown files Detect military IoT network of malicious threats
	Einstein Project (" National Cyberspace Security Protection System " NCPS)	Behavior detection capabilities based on machine learning (LRA)
	Social media manipulate	Processing and analysis of social media big data Natural language processing technology, image and emotion recognition technology Improve the scale and speed of collecting and processing social media data.
	Cyberspace Electronic Warfare	"Electronic Warfare Planning and Management Tool"(EWPMT) Machine learning upgrade and improvement

USArmy carries out independent networks, distributed processing in heterogeneous environments and application of machine learning systems research with machine learning. Through small samples, autonomous networks, fog edge computing, and cognitive science research we try to integrate the latest research breakthroughs in machine learning into the various tactical environments of the US military. US Navy's use of AI naval information respond to electronic attack project. Detect, identify, sense and defend are against new possible attacks. The U.S. Air Force's "Quantum Project", autonomous cluster flexible networks, machine learning wingmen and the development of sixth-generation aircraft have greatly increased its combat power. It has a leading advantage in terms of function construction and equipment R&D deployment. US Marines carry out the depth of reinforcement learning collaborative information warfare, the machine learning auxiliary aid cognitive decision making is integrated into its mandate to assess and weapon systems in. The US Space Force has carried out space-based data analysis and processing, and deployed machine learning platforms on supercomputing and cloud platforms to improve its ability to process massive amounts of real-time data. In particular, a series of plans of the US cyber forces cover multiple dimensions such as cyber threat defense, military IoT defense, machine learning behavior detection, social network data analysis, and cyber electronic warfare. Significantly they enhance its network attack and defense, social network behavior control and electronic warfare response capabilities (Table 3). Current ML works are applied in "back office" processes rather than on the front-line, etc.

## 6. Future trend, enlightenment and suggestions

Troops are taking machine learning technology into their land-based, sea-based, air-based, space-based and network space platforms system from different countries around the world. Our study summarizes the machine learning which plays a role in critical military applications in future operations, intelligence, networking, logistics, identification, health, trend data and key areas(Figure 1).

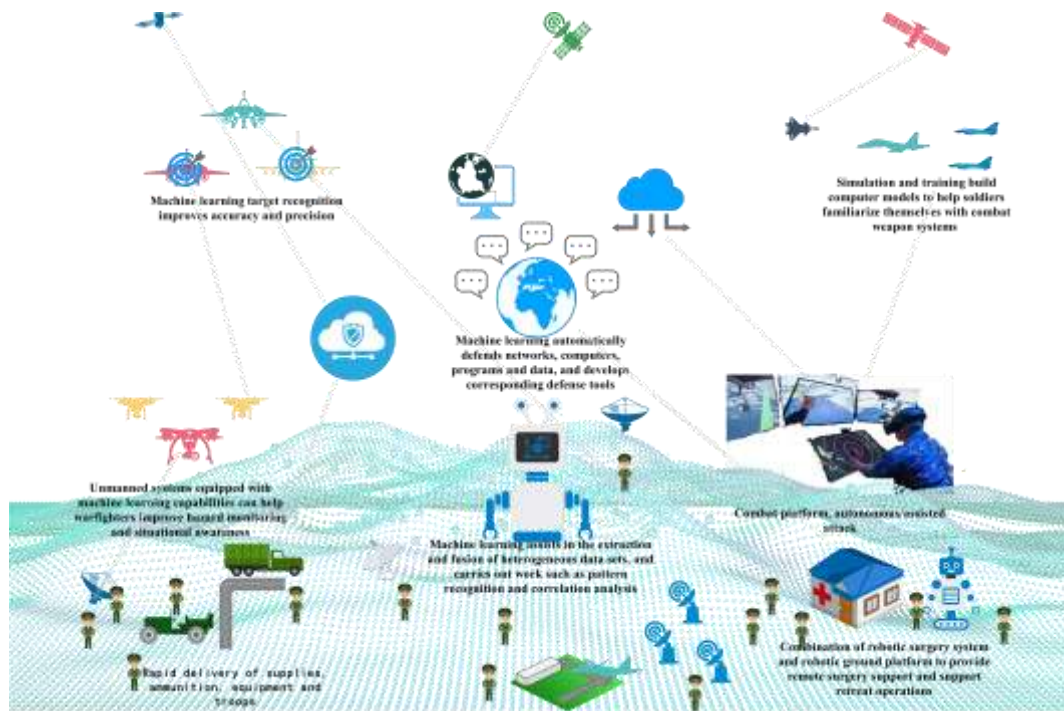


Figure 1: Machine learning focus to play the role of critical military applications in the future areas

## 7. Conclusion

With a new round of scientific and technological revolution, industrial revolution and industrial revolution led by the information revolution are intertwined evolving and have triggered a new round of military intelligence revolution. Machine learning has become one of the critical directions for countries to promote military modernization, and the global arms race in the field of artificial intelligence is becoming increasingly fierce. The militaries of different countries around the world are integrating machine learning technology into their weapons and other systems on land-based, sea-based, air-based, and space-based platforms. In addition, our research summarizes the potential of machine learning in multiple vital areas such as future operations, intelligence, network, logistics, identification, medical treatment, situation, and data. We have researched tracking and evaluation of global military machine learning technology development. We summarized the development status of global military machine learning technology, analyzed its main technology development path, studied its development trend, analyzed typical military application cases and development prospects of global military machine learning technology, and put forward potential.

\*The two authors contributed equally to this paper.

## Acknowledgements

We would like to thank the reviewer and editors for their professional advice, which has been of great help to our article adjustment and further research. This study was funded by the National Natural Science Foundation of China (Grant number 61862008, U1636208, 61902013), Shijiazhuang Science and Technology Research and Development Plan (Grant No. 201130351A) and Beihang Youth Top Talent Support Program (Grant No. YWF-20-BJ-J-1038).

## References

- Bastian, N. D. (2020). "Building the Army's Artificial Intelligence Workforce". *The Cyber Defense Review*, Vol 5, No. 2, pp 59–64.
- Hoadley, D. S. and Lucas, N. J. (2018). "Artificial intelligence and national security".
- Jian, Y. and Xianghua, B. (2020). "Research on the Dynamics of Artificial Intelligence Technology in the US Army". *Aerospace Electronic Warfare*, Vol 36, No. 1, pp 6–10.
- Mattis, J. (2018a). "Summary of the 2018 national defense strategy of the United States of America". Mattis, J. (2018b). "Summary of the 2018 national defense strategy of the United States of America".
- Mori, S. (2018). "US defense innovation and artificial intelligence". *Asia-Pacific Review*, Vol 25, No. 2, pp 16–44.

**Long Chen and Jianguo Chen**

- Tarraf, D. C., Shelton, W., Parker, E., Alkire, B., Gehlhaus, D., Grana, J., Levedahl, A., Leveille, J., Mondschein, J., Ryseff, J., et al. (2019). "The Department of Defense Posture for Artificial Intelligence".
- TASS (2020). "AI use in controlling weaponry in future will largely determine battle outcome — Putin".
- Xun, Z. (2020). "A comparative study on military applications of artificial intelligence in the United States and Russia". *National Defense Technology Industry*, , No. 01, pp 55–63.
- Zhimin, S. Z., Feifei, W., and Moon (2020). "Application progress of artificial intelligence in military confrontation". *JJ. Journal of Engineering Science*, Vol 2020, No. 09, pp 1106–1124.
- Эльяс Касми (2020). "В российских министерствах и госведомствах появится ИИ для поиска преступников и работы с документами"

# Global Social Network Warfare on Public Opinion

Long Chen<sup>1,2</sup> and Jianguo Chen<sup>31</sup>

<sup>1</sup>Beijing Key Laboratory of Network Technology, Beihang University, Beijing, China

<sup>2</sup>Innovation Technology Research Institute, Beijing Topsec Network Security Technology Co Ltd, China

<sup>3</sup>Hebei Seismological Station, Earthquake Administration of Hebei Province, Shijiazhuang, China

[zhuanjiatuijian@126.com](mailto:zhuanjiatuijian@126.com)

DOI: 10.34190/EWS.21.093

**Abstract:** Information warfare can be divided into two types: technical and psychological information warfare. The scope of cyberspace weapons is expanding from the physical network domain to the cognitive information domain. Technologies such as network penetration, public opinion guidance and attacks, cognitive intervention and control will become the main development directions. Controlling public opinion and controlling audiences will become the cyberspace cognitive domain. The development goal of the weapon. In recent years, emerging network media such as social networks and mobile communication networks have played an important organizational and planning role in a series of significant events, which are likely to cause severe threats to national security and even the international community's stability. We conclude that social public opinion weapons are mainly divided into six categories: Bot, Botnet, Troll, Manipulate real people and events, Cyborg, and Hacked or stolen. Since social network warfare is a brand new war situation in the context of great powers, in social media, the confrontation of camps can be observed. States use social media platforms to penetrate and media war, and its Internet space monitor and build defenses. A digital wall has been placed horizontally on the boundary of the virtual space. In recent years, as the trend toward weaponization of social media has become increasingly apparent, military powers such as the United Kingdom, the United States, and Russia have taken the initiative to take the lead in the field of social media. All countries continue to strengthen the research on fundamental cognitive theories. Many basic research projects that integrate information, biology, network, and cognition have been launched one after another. The combat practice shows that the effectiveness of social media even exceeds some traditional combat methods. With the prominent role of social media in modern warfare, its combat use has become increasingly widespread and has gradually become a force multiplier in modern warfare. At present, the deployment of weapons directed by social network users in significant countries is mainly focused on incident and public opinion reconnaissance, sentiment analysis, and active intervention. In short, all countries are using social media to spread political propaganda and influence the digital information ecosystem. The technical means, scale, scope, and precision of social media weapons have been continuously improved. It is gaining momentum to reshape the cyberspace security pattern of various countries fundamentally.

**Keywords:** social network, global warfare, public opinion

---

## 1. Introduction

In recent years, as the weaponization trend of social media has become increasingly prominent, military powers such as the United Kingdom, the United States, and Russia have taken the initiative to take measures such as setting up specialized agencies, issuing relevant regulations, and increasing technological research and development, in an attempt to take control of the social media field. In 2019, Oxford University researchers released a report (Bradshaw & Howard, 2019) showing that more than 56 countries carry out cyber military activities on Facebook. All countries continue to strengthen the research on fundamental cognitive theories. Among them, the US Department of Defense has successfully launched many basic research projects on the intersection of information, biology, network, and cognition, focusing on the study of neurocognitive functions and neural tissues to explore human information acquisition's cognitive mechanism.

In 2007, the National Research Council of the US Academy of Sciences published the report "Army Network Science, Technology and Experimental Center Policy" (Council, 2007), setting out the Army's priority investment areas in the field of network infrastructure. And the use of social networks followed by humans. This has become an important area for research.

US Defense Advanced Research Projects Agency 2011 published the "Social Media Strategic Communications (Social Media in Strategic Communication, SMISC) plan" (Hsieh, 2015), intended to enhance the ability of US forces

---

<sup>1</sup> Corresponding author: [zhuanjiatuijian@126.com](mailto:zhuanjiatuijian@126.com) (Jianguo Chen)

to carry out professional guidance of public opinion. Through careful design, coordination, coherence, and management of these social activities' influence, its effectiveness will be significantly improved.

In Russia IRA (Internet Research Agency) Internet research institutions (Agentstvo Internet- Issledovaniy, also known as Glavset) (Lapowsky, 2017), and is referred to in the Russian city of well-known Internet companies Olgino of Trolls, is a representative of the Russian business and politics Interest in Russian companies engaged in social network influence tasks (Prier, 2017).

According to research conducted by Ali Murat Kirik of the University of Marmara, Turkey, social media has become one of the essential warfare tools and is known as a new generation of war methods, including deceptive news, diplomacy, law, and foreign election interference. The increasing analyze the number of perceptual management and social engineering activities in social media proves its use as a war tool. Social media can shape ideas and redesign society. It can increase sensitivity and strengthen the social response. Anything can become a reality through social media, and society can become part of this simulated or artificial world. Social media can manipulate reality to weaken the administration, society, military, or economy of a country (Aksüt, 2020).

In recent years, the US military plans to analyze worldwide more than 3500 billion social media posts to help track the mass movement's evolution. With 60 languages to collect screened from at least 100 countries, two info million users understand the "group expression pattern" better. The project will research the comments, metadata, location and hometown identifiers of messages including user names.

In 2019, Oxford University researchers released a report (Bradshaw & Howard, 2019) that more than 56 countries carry out cyber military activities on Facebook. Researchers found that due to its market size and influence, it can spread political news and information and form groups and topics, which can also be widely used as a social network weapon. All countries are using social media to spread political propaganda and influence the digital information ecosystem. The scale, scope, and precision of social media weapons have been continuously improved. It is gaining momentum the cyberspace security pattern of various countries to fundamentally re-shape.

From 1st January 2020, the US Carnegie Mellon University Researchers studied 2 Tweets discuss new crown epidemic and related issues more than 100 million, according to the researchers, in discussions on the virus tweets, nearly half of them are sent by robots (Young, 2020).

## **2. Types of weapon arsenal on social public opinion**

Social public opinion information warfare includes technical and psychological information warfare. The scope of cyberspace weapons is expanding from the physical network domain to the cognitive information domain. Technologies such as network penetration, public opinion guidance, and attacks, cognitive intervention and control will become the main development directions. Controlling public opinion and controlling audiences will become the cyberspace cognitive domain—weapon development goals (Shen et al, 2015). In recent years, emerging network media such as social networks and mobile communication networks have played an important organizational and planning role in a series of significant events, which are likely to cause severe threats to national security and even the international community's stability. As the social public opinion weapon arsenal's core equipment, social robots create content and messages, which are injected into online social platforms, and read and spread by others (Figure 1). Deceptive news stories, invasion and release of private communications, fabricated incidents, statements or results, and the spread of fear have been used to influence the target country's politics, military, and economy (Wang). In the field of social media, the confrontation of the camp can be observed. Western use of social media platforms and media penetration war, and its Internet space monitor and build defenses. A digital wall has been placed horizontally on the virtual space boundary (Qianye, 2019).

Social network arsenals are summarized (Table 1).

**Table 1:** Summary of social network arsenal

<b>Classification</b>	<b>Definition</b>	<b>Judge</b>	<b>Means</b>
Bot	An automated program used to participate in social media. Bots are automated social media accounts operated by algorithms (DFRLab, 2018)	There are no personalized posts, comments, answers, or interactions with other users' online posts	Ostensibly expand the visibility of an individual or organization Influence election



Classification	Definition	Judge	Means
			Manipulate financial markets Amplify phishing attacks Spread spam Destroy freedom of speech
Botnet	Refers to those who are motivated by commercial interests to achieve improper purposes.(Benevenuto et al, 2010; Bouguessa, 2011; Halpin & Blanco, 2012; Hayati et al, 2010; Langbehn et al, 2010; Liu, 2012; Raykar & Yu, 2012; Wang et al, 2011)	Network botnet identification uses Web information mining technology in the current network environment (Russell, 2013) to define high-discrimination features and behaviour patterns to discover hidden network botnet. The network botnet can also be understood as an outlier among all network users (Jiang et al, 2010).	Internet writer Manufacturing and using false information Re-comment and big V means Download and play data fraud Anti-discrimination ad Commercial and political robots
Troll	People who deliberately initiate online conflicts or offend other users, distract and create divisions in online communities or social networks by publishing inflammatory or non-themed posts.	The difference between a troll and a robot is that a troll is a real user, and the robot is automatically set. The two types are mutually exclusive and do not overlap	normal Profanity Trolling Derogatory Hate speech (Aggarwal et al, 2020)
Manipulate real people and events	Try to connect with individuals and mobilize them to take action in the real world	Influencing unsuspecting people can range from forwarding or disseminating propaganda, to hiring false intelligence personnel online to persuade real people taking action.	"Useful Idiot" "Fellow Traveler" "Secret Agent"(Clifton, 2018)
Cyborg	Integration of automation and manual control	Use auxiliary tools while being controlled by people	Mass forwarding, fast reply and ultra-high frequency posting
Hacked or stolen	The hacked or stolen account	Highlights the Internet's propaganda through more traditional forms of cyber attacks	"False" high-profile account is a network of strategic forces to disseminate pro-government propaganda or revoke the rightful owner by reviewing account access

## 2.1 Bot

A social media robot (Bot) is an automated program to participate in social media. A bot is an automated social media account operated by an algorithm. These robots exhibit partially or utterly autonomous behaviour and are usually designed to mimic human users. Robots' key indicators are anonymity, high activity, and amplification of specific users, topics, and tags (DFRLab, 2018).

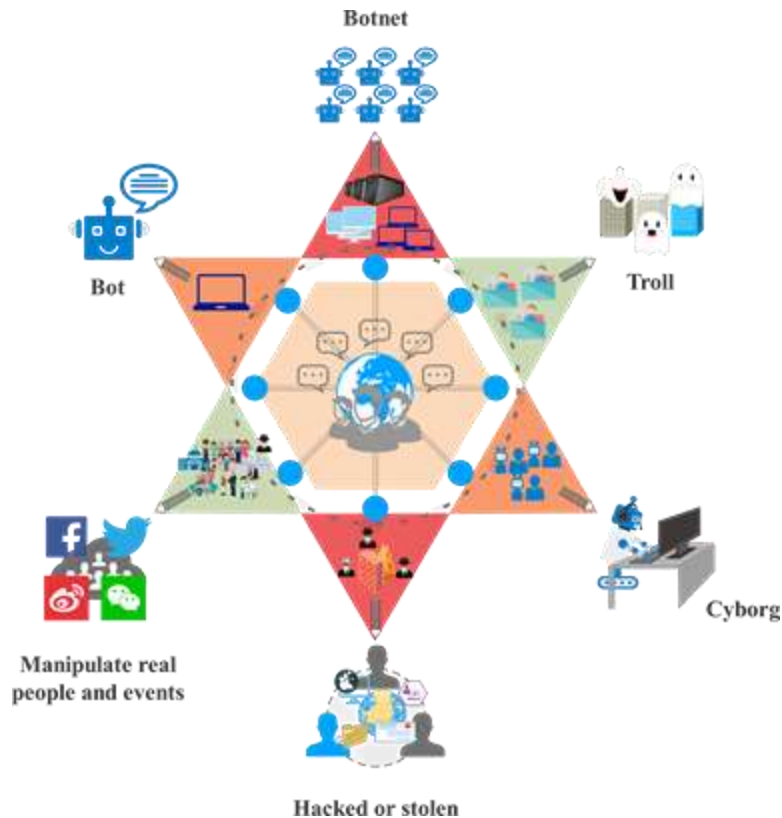
Automated media accounts on social networking sites allow social media users to artificially enlarge and increase dissemination, or the "toxicity" of network content. Russian-backed agents have used these automated accounts and are committed to developing and improving their robotic capabilities to spread false information faster and deeper across the social media field. 2018 Nian 1 Yue, Twitter disclosed its security personnel assessment that there are more than 5 Wan automatic accounts related to Russia in Twitter post-election-related content during the US presidential campaign (Blog, 2018).

A large number of social media bots are malicious bots disguised as human users. Malicious social media bots can be used to influence politics, advertising, access, networks, and media for multiple purposes (DFRLab, 2018).

## 2.2 Botnet

The botnet is a network of robot accounts managed by the same person or group. The so-called “water army” refers to the Army of fans lurking on the Internet social platforms to bluff for specific users or topics. Most of these paid senders and commentators of false information are employed by relevant agencies or competitors in the target country.

The network of water army is those driven by commercial interests, to achieve a network such as the impact of public opinion, disrupting the network environment and another improper purpose, by manipulating the software robot or botnet accounts, made on the Internet, spreading false information, comments and spam comments (Benevenuto et al, 2010; Bouguessa, 2011; Halpin & Blanco, 2012; Hayati et al, 2010; Langbehn et al, 2010; Liu, 2012; Raykar & Yu, 2012; Wang et al, 2011).



**Figure 1:** Schematic diagram of the social network weapon arsenal

Network botnet recognition uses Web information mining technology in the current network environment (Russell, 2013) to define high-discrimination characteristics and behaviour patterns to discover hidden network botnet. The network botnet can also be understood as an outlier among all network users (Jiang et al, 2010). Because of the social networking media robot clearance and enhanced detection method, Botnet network herdsman becomes more carefully, so that a single robot is more difficult to detect (DFRLab, 2018). However, evidence such as speech patterns, identical posts, date and time of creation, behavioural relevance, location relevance, and account name patterns can still be used to distinguish the botnet (DFRLab, 2018). The botnet’s methods are divided into network writers, making and using false information, reposting likes and oversized V methods, downloading and playing data falsification, anti-serial slander advertising, commercial and political robots, etc.

## 2.3 Troll

Troll deliberately initiates conflict or offends other users online, distracting and dividing people in the online community or social network or posting provocative topic posts. Their goal is to irritate others, react emotionally, and disrupt discussions. The difference between a troll and a robot is that a troll is a real user, while the robot is automatically set. The two types are mutually exclusive and do not overlap. Trolling as a behaviour, not just limited to troll would do.

The troll arsenal is divided into five categories: normal, profanity, trolling, derogatory, and hate speech (Aggarwal et al, 2020). A network troll is a keyboard player who posts provocative, offensive, harassing or misleading information on the Internet by a real person who tries to provoke reactions from other users on social media.

In recent years, Russia's use of troll behaviour is more and more exposed. Entity supported by Kremlin has specialized in operating a large number political funding troll troops over the years. Investment construction of large-scale industrialized "troll farm" and the Russian advanced in the global information goals.

The NATO Center for Strategic Communications of Excellence was commissioned to research trolls in hybrid warfare. This research focuses on discussions surrounding Ukraine and Russia's conflict and outlines the influence techniques used by cyber trolls, including improper use. The slander and attack, the use of irony and satire, the peddling of conspiracy theories, the use of young and attractive male and female avatars, the transfer of topics to others, the posting of misleading information on sources such as Wikipedia, and the emphasis on social issues e.g. the gap between rich and poor, without source or verification, provides a lot of indigestible data (Chen, 2016; Svetoka, 2016).

## **2.4 The manipulation of real people and events**

While creating deliberate disinformation, trolls, we try to establish contact with individuals and mobilize them to take action in the real world. Russian spies hired bogus intelligence agents online to attract a wide range of real people to take action from influencing unsuspecting people to forwarding or disseminating propaganda to persuading someone to stage a real-world protest.

In 2017 Clint · Watts (Clint Watts) outlined three different types of Russian influence operations potential organizational penetration of target (Watts, 2017). One category of "useful idiots" refers to unsuspecting Americans being used to further expand their propaganda in Russia. "Family travelers" ideologically sympathize with Russia's anti-Western elements to take action on their own. "Secret agent " refers to being actively manipulated to carry out illegal or secret operations on behalf of the Russian government (CliGon, 2018).

## **2.5 Cyborg**

The robot account, which combines automation and manual control, is another weapon account type. It is controlled by humans while using auxiliary tools to realize large-scale forwarding, quick reply, and ultra-high-frequency posting accounts in a short time -this kind of account is also called "Cyborg (semi-robot)," which is usually with a strong sense of purpose.

## **2.6 Hacked or stolen**

Hacked or stolen accounts. Although these accounts are not robot accounts in nature, "deception" high-profile accounts are strategic cyber forces to spread pro-government propaganda or revoke access to legitimate owners' accounts through censorship. A small number of national cyber forces have begun to use stolen or hacked accounts as part of their activities, highlighting the Internet's use of more traditional forms of cyberattacks for propaganda. Also, not all of the accounts used by the military are false.

## **3. Social public opinion combat effect**

The combat practices of Ukraine, Iraq, and Syria have shown that social media may produce effects that aircraft, tanks, and artillery cannot achieve. In a sense, its effectiveness even exceeds some traditional combat methods (Chen & Xia, 2016). With the prominent role of social media in modern warfare, its combat use has become increasingly widespread and has gradually become a force multiplier in modern warfare.

### **3.1 Use social media to gather intelligence information**

Various national military and political departments, organizations, and individuals conduct background analysis of messages, photos, and videos on Twitter, Facebook, YouTube, Weibo, and WeChat to identify the details of the publisher's activities and implement unique intelligence collection.

Superpowers support the release of cyber intelligence perception weapon arsenal capabilities through the framework of global intelligence monitoring projects. Especially Edward Snowden (Gellman, 2020) in 2013, the US National Security Agency (NSA) had a global intelligence, surveillance and disclosure of its various global intelligence surveillance project intensified. Large-scale surveillance is considered by some countries as one of the means to combat terrorism, prevent crime, prevent social unrest, and protect national security.

June in 2015, a commander "Islamic State", exposing its position in the building because of the organization on social network released a self portrait. Then US warplanes were bombing killed.

Since 2017 the attempted attacks on electronic devices have effectively diverged people's views on the government, media and public institutions in Ukraine, Bulgaria, Estonia, the United States, Germany, France and Austria. Because it is difficult to blame, the conduct of these actions will not expose the aggressor to political severe or military risks.

### **3.2 Influence public perception through social media**

Cyber forces use a variety of messaging and value strategies when communicating with users online. At the same time, the influence of public awareness on military operations is increasing. Social media can quickly push selected information to hundreds of millions of target audiences in the form of pictures and texts, which affects the target audience's views and attitudes towards the event.

### **3.3 Use social media to implement psychological deterrence**

In conflicts during the 21st century, the psychological level is as crucial as the physical level, and the creative use of psychological warfare can effectively offset the opponent's advantage on the physical battlefield. Social media has become an essential means through which extremist terrorist organizations gain popularity.

In recent years, the combat use of social media has shown a diversified trend. The "Islamic State" uses social media extensively to spread revolutionary ideas, recruit members and raise funds (Alava et al, 2017).

## **4. The social public opinion arsenal module evolution**

In terms of the cyber social weapon arsenal, organizations with military backgrounds in various countries have created cyber social public opinion cyber weapons. Through the cultivation of social robots (robot "big V"), they affect the information environment of the target country.

By creating a new type of strategic game in cyberspace and the sharpest social network weapon for security struggles, countries have realized the guidance of online public opinion, which mainly refers to the long-term confrontation of various communication channels surrounding network information, competing for the dominance of public opinion and then ideology. Social network behaviour guidance is a series of activities to change the attitude and behaviour of users or groups with specific intentions. People need to integrate different sources of information when making decisions, and in social media, the information environment may be distorted by the influence of cyberspace robot account weapons.

Since social network warfare is a new war situation under the background of the game of big powers, the current weapon deployments directed by social network users in major countries mainly focus on incident and public opinion reconnaissance, sentiment analysis, and active intervention:

The weapon means of the event and public opinion reconnaissance includes two continuous identification and tracking processes. Solutions to this task can be divided into four categories: topic-based model, text-based clustering, feature-based, and pattern-based.

The task of online public opinion warfare can divide sentiment analysis into sentiment polarity classification, sentiment change conditions, subjective and objective analysis, sentiment analysis combined with multiple features, and feature selection in sentiment analysis.

In terms of active intervention technology of network social weapons, it is mainly divided into monitoring and intervention. Such hostile forces in Twitter collect relevant information and extract the user's mood and feel a

sense of social robot to identify opinion leaders, influential users, cultivating big V social robot account intervention, guiding public opinion, and influencing political outcomes.

Our research on social network weapons mainly consists of the following modules (Table 2 ).

**Table 2:** Social network weapon module

Weapon module	Technical means
Interactive information characterization module	To achieve the problem of representation of interactive text, social text multi-level and multi-angle graph representation methods is used, including research on the extraction of multi-dimensional information. The interactive information semantic variable granularity representation realizes social text lexical, sentence and paragraph level multi-level semantic representation.
Behaviour Information Representation Module	Construct a multi-dimensional representation method of user operation behaviors of four types of social network likes: favorites, comments, and forwards, realize the characterization of each dimension, construct a text topic extraction method based on topic models, and construct a text source authority and text emotional tendency Characterization algorithm, the expression method of text popularity, and the calculation method of poor behavior.
Character portrait information representation module	Realize the characterization problem of portrait information, and construct the description method of the user's basic attributes, personality characteristics, interest characteristics, and emotional characteristics. We use data mining methods to obtain the user's basic attributes, use the Big Five Model theory to analyze the text posted by the user to obtain the user's personality label, use machine learning methods to train the text feature and the user's personality association model, build a key Word extraction, and sentiment analysis are used to obtain user interest weights to construct user sentiment representations based on short texts.
User recognition model construction	Implement depicts the problem of social network users cognitive processes and builds relationships between the user and the user accepts portraits of information to build user awareness and user behavior model to describe the user cognitive processes.
User behavior guidance mechanism construction	The social network behaviour guidance mechanism realizes the guidance of the user's original behaviour to the desired behaviour. The guidance mechanism's construction includes three aspects: the conversion of the guidance intention to the desired user behaviour, the cognitive reasoning function based on the convex function to find the optimal semantic difference, and the template-based guidance text generation.
Homogeneous community discovery	The commonalities in homogeneous communities are called homogenous attributes. The homogeneity of the property is the basis for reaching a consensus and forming the convergence of group behaviour. Homogeneous community discovery requires the adoption of behaviour-driven social network structure timing models and community discovery techniques based on consistent attributes.
Behavioural collaboration module of homogeneous groups	Realize the convergent trigger problem of social network group behavior, build a guided communication model based on guiding text, and a group interaction model based on sentiment analysis. Construct a multi-source social information dissemination model, and realize the prediction of the scale of the guidance text in the social network; design the information-oriented communication maximization method to realize the homogeneity of the guidance text in the dissemination process full coverage of the community.

## 5. Conclusion

In recent years, emerging network media such as social networks and mobile communication networks have played an important organizational and planning role in a series of major events, which are likely to cause extremely serious threats to national security and even the international community's stability. We conclude that social public opinion weapons are mainly divided into Bot, Botnet, Troll, Manipulate real people and events, Cyborg and Hacked or stolen. Because social networking is seen as a new weapon in the great powers' battle ground arsenal, it has been the subject of widespread military research. All countries continue to strengthen the research on basic cognitive theories. With the prominent role of social media in modern warfare, its combat use has become increasingly widespread and has gradually become a force multiplier in modern warfare. At present, the deployment of weapons directed by social network users in major countries is mainly focused on incident and public opinion reconnaissance, sentiment analysis, and active intervention. In short, all countries are using social media to spread political propaganda and influence the digital information ecosystem. The technical means, scale, scope, and precision of social media weapons have been continuously improved. It is gaining momentum towards fundamentally reshaping cyberspace.

\*The two authors contributed equally to this paper.

## Acknowledgements

We would like to thank the reviewer and editors for their professional advice, which has been of great help to our article adjustment and further research. This study was funded by the National Natural Science Foundation of China (Grant number 61862008, U1636208, 61902013), Shijiazhuang Science and Technology Research and Development Plan (Grant No. 201130351A) and Beihang Youth Top Talent Support Program (Grant No. YWF-20-BJ-J-1038).

## References

- Aggarwal, K., Bamdev, P., Mahata, D., Shah, RR & Kumaraguru, P. (2020) Trawling for Trolling: A Dataset. *arXiv preprint arXiv:2008.00525*.
- Aksüt, F. (2020) *Social media evolves to warfare tool: Expert*, 2020. Available online: <https://www.aacom.tr/en/science-technology/social-media-evolves-to-warfare-tool-expert/1818953>
- Alava, S., Frau-Meigs, D. & Hassan, G. (2017) *Youth and violent extremism on social media: mapping the research* UNESCO Publishing.
- Benevenuto, F., Magno, G., Rodrigues, T. & Almeida, V. (2010) Detecting spammers on Twitter, *7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference, CEAS 2010*.
- Blog, TPP (2018) Update on Twitter's review of the 2016 US election. January 19. Available online: [https://blog.twitter.com/en\\_us/topics/company/2018/2016-election-update.html](https://blog.twitter.com/en_us/topics/company/2018/2016-election-update.html)
- Bouguessa, M. (2011) An unsupervised approach for identifying spammers in social networks, *Proceedings- International Conference on Tools with Artificial Intelligence, ICTAI*.
- Bradshaw, S. & Howard, PN (2019) *The global disinformation order: 2019 global inventory of organized social media manipulation* Project on Computational Propaganda.
- Chen, A. (2016) The Real Paranoia-Inducing Purpose of Russian Hacks, *The New Yorker*. 2016
- Chen, H. & Xia, Y. (2016) Social media war: a new dimension of war in the information age. *Grand Garden of Science* (2016/03), 38-40.
- CliGon, D. (2018) A Murder Plot, a Twitter Mob and the Strange Unmasking of a Pro-Kremlin Troll, *Mother Jones*. June 5 2018
- Council, NR (2007) *Strategy for an Army Center for Network Science, Technology, and Experimentation* National Academies Press.
- DFRLab (2018) *TrollTracker: Bots, Botnets, and Trolls*, 2018. Available online: <https://medium.com/dfrlab/trolltracker-bots-botnets-and-trolls-31d2bdf4c13>
- Gellman, B. (2020) *Dark mirror: Edward Snowden and the American Surveillance State*. New York: Penguin Press.
- Halpin, H. & Blanco, R. (2012) Machine-learning for spammer detection in crowd-sourcing, *AAAI Workshop- Technical Report*.
- Hao, K. (2020) *Nearly half of Twitter accounts pushing to reopen America may be bots*, 2020. Available online: <https://www.technologyreview.com/2020/05/21/1002105/covid-bot-twitter-accounts-push-to-reopen-america/>
- Hayati, P., Chai, K., Potdar, V. & Talevski, A. (2010) *Behaviour-based web spambot detection by utilizing action time and action frequency*. 2010.
- Hsieh, M. (2015) Social Media in Strategic Communication (SMISC). *Program Information*, 11.
- Huo, G. (2017) *The social mobilization role of social media in the "Arab Spring"* Master. China Foreign Affairs University.
- Jiang, F., Du, JW, Sui, YF & Cao, CG (2010) Outlier detection based on boundary and distance. *Tien Tzu Hsueh Pao/Acta Electronica Sinica*, 38(3), 700-705.
- Langbehn, H., Ricci, S., Gonçalves, M., Almeida, J., Pappa, G. & Benevenuto, F. (2010) A multi-view approach for detecting noncooperative users in online video sharing systems. *J. of Inf. and Data Management*, 1(3), 313-328.
- Lapowsky, I. (2017) *Facebook May Have More Russian Troll Farms to Worry About*, 2017. Available online: <https://www.wired.com/story/facebook-may-have-more-russian-troll-farms-to-worry-about>
- Liu, QW (2012) Web spammers' detection and prevention. *Press Outpost*, 6, 021.
- Mangan, D. & Calia, M. (2018) *Special counsel Mueller: Russians conducted information warfare against US to help Trump win*, 2018. Available online: <https://www.cnbc.com/2018/02/16/russians-indicted-in-special-counsel-robert-muellers-probe.html>
- Prier, J. (2017) Commanding the Trend: Social Media as Information Warfare. *Strategic Studies Quarterly*, 11(4), 50-85.
- Programs, USSDB o. II (2015) *Everything you wanted to know about trolls but were afraid to ask* 2015. Available online: <https://share.america.gov/trolls-everything-you-wanted-to-know>
- Qianye, Z. (2019) *The global social media market in the age of imperialism*, 2019. Available online: <https://zhuannlan.zhihu.com/p/86096911>
- Raykar, VC & Yu, S. (2012) Eliminating spammers and ranking annotators for crowdsourced labeling tasks. *Journal of Machine Learning Research*, 13, 491-518.
- Robertson, A. (2018) *Facebook suspends 273 accounts and pages linked to Russian misinformation agency*. *theverge.com*. Retrieved, 2018. Available online: <https://www.theverge.com/2018/4/3/17194518/facebook-suspends-russian-internet-research-agency-pages-accounts-instagram>

**Long Chen and Jianguo Chen**

- Russell, MA (2013) *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*.
- Seddon, M. (2014) *Documents Show How Russia's Troll Army Hit America*, 2014. Available online: <https://www.buzzfeednews.com/article/maxseddon/documents-show-how-russias-troll-army-hit-america>
- Shen, X., Wu, J. & Deng, Q. (2015) Development Trends of Cyberspace Weapon Systems in US Army. *Journal of Equipment Academy*, 26(06), 70-73.
- Svetoka, S. (2016) *Social Media as a Tool of Hybrid Warfare*, 2016. Available online: <https://www.stratcomc.oe.org/social-media-tool-hybrid-warfare>
- Wang, G., Xie, S., Liu, B. & Yu, PS (2011) Review graph based online store review spammer detection, *Proceedings IEEE International Conference on Data Mining, ICDM*.
- Wang, X. *Web strategy*, 1 edition. Fudan University Press.
- Watts, C. (2017) *Hearing before the Senate Armed Services Committee*, 2017. Available online: <https://www.fjiri.org/wp-content/uploads/2017/04/Watts-Testimony-Senate-Arrried-Services-email-distro-Final.pdf>
- Young, VA (2020) *Nearly Half of the Twitter Accounts Discussing 'Reopening America' May Be Bots*, 2020 Available online: <https://www.scs.cmu.edu/news/nearly-half-twitter-accounts-discussing-reopening-america-maybe-bots>

# Serious Games for Cyber Security: Elicitation and Analysis of End-User Preferences and Organisational Needs

Sabarathinam Chockalingam, Coralie Esnoul, John Eidar Simensen and Fabien Sechi  
Institute for Energy Technology, Halden, Norway

[Sabarathinam.Chockalingam@ife.no](mailto:Sabarathinam.Chockalingam@ife.no)

[Coralie.Esnoul@ife.no](mailto:Coralie.Esnoul@ife.no)

[John.Eidar.Simensen@ife.no](mailto:John.Eidar.Simensen@ife.no)

[Fabien.Sechi@ife.no](mailto:Fabien.Sechi@ife.no)

DOI: 10.34190/EWS.21.043

**Abstract:** Digitalisation is more actual than ever and even forced by the Covid-19 pandemic for many. The evolution of technology enables everyone and everything to be connected. This is one of the reasons why cyber security is important to society as it makes the large majority vulnerable to cyber-attacks. Cyber-attacks not only impact confidentiality, integrity and availability of information but also can cause physical damage like Stuxnet. Notably, humans are considered the weakest link in cyber security. Training plays an important role in strengthening the weakest link. A survey was conducted with the aim of developing a serious game for cyber security training where we found that current cyber security trainings are not effective in practice. The survey results showed that the conventional training method is both widely used and at the same time considered the least preferred training method. On the other hand, the game-based training method seems to be the least used training method, but this seems to be one of the most preferred training methods. Existing serious games in cyber security are “generic” as they do not seem to neither consider end-user preferences nor can be tailored to the specific and varying needs of an organisation. Therefore, a survey was conducted in an organisation to elicit end-user preferences. This was complemented with interviews of key management personnel to gather organisational needs. Based on the analysis of survey and interview results, a set of requirements are provided for developing a serious game for cyber security training in a specific organisation.

**Keywords:** cyber security, organisational needs, serious games, training, user requirements

---

## 1. Introduction

The Covid-19 pandemic has created new opportunities for cyber-criminals due to changes in working practices of organisations. This has also resulted in increasing number of cyber-attacks (Lallie et al., 2020). A survey shows that 60% of Small and Medium-sized Enterprises (SMEs) that experience a cyber-attack goes out of the business within six months (Ponsard and Grandclaoudon, 2018). In cyber security, humans are often referred to as the weakest link (Pfleeger et al., 2014). This is also evident from cyber-attacks towards a German steel mill (Lee et al., 2014) and the Norwegian parliament (BBC-News, 2020). In addition, phishing is the most common starting point of cyber-attacks launched during the Covid-19 pandemic (Lallie et al., 2020). Training humans play an important role in effectively dealing with such cyber-attacks (Thomas, 2018). Typically, a cyber security training utilises one of the following seven training methods: (i) conventional (or paper-based) method (example: poster), (ii) instructor-led (or classroom-based) method (example: lecture on basics of cyber security), (iii) online-based method (example: email, online discussion), (iv) game-based method (example: serious game to raise employee awareness in cyber security), (v) video-based method (example: cyber security awareness video on phishing), (vi) simulation-based method (example: attack simulators to raise employee awareness in cyber security) and (vii) event-based method (example: lunch seminar) (Abawajy, 2014, Ghazvini and Shukur, 2017). (Ghazvini and Shukur, 2017) highlighted that the game-based training method possess the key training success factors like challenge, fun and motivation whereas the conventional method lacks these training success factors. Therefore, we intend to develop a serious game for cyber security trainings in an organisation as it supports both the key training success factors and end-user preferences on training methods.

The primary objective of a serious game is training rather than fun (Wattanasoontorn et al., 2013). Furthermore, serious games are useful especially in domains like cyber security as it is difficult to conduct trainings in a real environment involving real systems. However, the use of realistic environments with relevant storylines enables trainees to apply lessons learnt from the serious game in real-life situations. Serious games are mainly used for training purposes in different domains such as emergency management (Van Ruijven, 2011, Chittaro and Ranon, 2009) and medical (Wattanasoontorn et al., 2013, Graafland et al., 2012). Furthermore, serious games are also used in cyber security which will be detailed in Section 2.



The development of a serious game typically follows the main phases of Design Science Research (DSR) method: (i) problem identification, (ii) objectives of a solution, (iii) design and development and (iv) evaluation (Offermann et al., 2009, Ávila-Pesántez et al., 2017). Serious games used for cyber security trainings are generic as they follow one-size-fits-all approach without considering preferences from end-users and the needs of the organisation. The one-size-fits-all approach can result in unwanted game features and lack of storylines relevant for a specific organisation. This in turn make cyber security trainings ineffective in practice. Our research aims to fill this gap by addressing the research question (RQ): “How could we elicit and analyse end-user preferences and organisational needs to develop a serious game for cyber security trainings in an organisation?”. This RQ mainly corresponds to the second phase (objectives of a solution) of DSR method. This paper is structured as follows: In Section 2, we analyse existing serious games in cyber security. In Section 3, we describe the study method, followed by the results in Section 4. Section 5 discusses implications, generalisability, and limitations of this study. Section 6 presents conclusions and future research directions.

## **2. Related work**

This section analyses serious games in cyber security using different criteria and identify important usage patterns. Furthermore, we also highlight the key issue which we use as a basis to address in this study. We relied on existing reviews of serious games in cyber security (Tioh et al., 2017, Alotaibi et al., 2016) and complemented it with a publicly available repository on security games (Shostack, 2018) to identify serious games in cyber security. Although not all existing serious games are considered, the sources which we relied on should provide a good coverage. For instance, the repository is constantly updated with the last update in January 2021 and maintained by an expert game designer in security. In addition, we compared the identified games using four different criteria: (i) learning emphasis, (ii) target group, (iii) digital/non-digital game and (iv) single/multi-player game. This comparison can provide a basis to develop effective serious games in cyber security in the future. For instance, this analysis shows which topics in cyber security were mainly addressed in existing serious games in cyber security and point out topics that need more attention in the future.

We identified 46 serious games in cyber security which were released between 2004 to 2020. 17 out of 46 serious games can only be played digitally, 28 out of 46 serious games can only be played physically and only one out of 46 serious games can be played both digitally and physically. Furthermore, 15 out of 46 serious games were single-player games, 24 out of 46 serious games were multi-player games and 7 out of 46 serious games had both single-player and multi-player options. The following cyber security areas/problems were mainly addressed in the identified serious games: (i) network security, (ii) cyber security terminologies/concepts (teaching), (iii) incident response, (iv) social engineering (phishing) and (v) online security. On the other hand, there were very little or no attention to serious games that focus on: (i) cyber security of Operational Technology (OT) environment, (ii) cyber security of SMEs, (iii) social engineering attacks except phishing and (iv) cyber-attacks initiated through physical means.

Decision-makers, developers, designers, ICT users (employees), incident-responders and security professionals were the target group in an organisation for serious games, whereas some serious games specifically targeted, children, teenagers, and students. Most of the serious games in cyber security that target security professionals, decision makers, developers, designers, and incident responders were table-top games, which acts as a tool to facilitate discussion or brainstorming. There were serious games which specifically addressed some of the common cyber security threats listed in well-known reports (Sfakianakis et al., 2019, Verizon, 2019). For instance, “Anti-Phishing Phil” addressed phishing. However, common cyber security threats like ransomware that were not addressed in the identified serious games in cyber security. Context is missing in the identified serious games in cyber security as they were generic which might not be effective when we use it for a specific organisation. This also led to lack of realistic storylines in these serious games which would help to engage the users. Therefore, the present study focuses on improving this aspect by gathering end-user preferences and organisational needs and then translating it into high-level requirements to develop a serious game in cyber security tailored to a specific organisation.

## **3. Method**

This section describes how we elicited and analysed end-user preferences and organisational needs. We chose questionnaire as the method for gathering end-user preferences as it supports the collection of data from a large number of employees in an organisation within a short period of time compared to other data collection methods like interviews and focus groups. This method is also practical and safe considering the Covid-19

pandemic during which this study was conducted. An excerpt of the questionnaire including the complete set of questions used to elicit end-user preferences can be found in Appendix A. The major objective of this questionnaire is to gather end-user preferences of cyber security training methods and game features. Firstly, a draft questionnaire was developed and then reviewed by two of the authors who were not primarily involved in the design of it. The questionnaire was updated based on the review inputs. The updated questionnaire was then validated with a Chief Information Security Officer (CISO) of an organisation. Finally, we sent the developed questionnaire to the employees of an organisation in which we intend to develop a game-based cyber security training. We utilised interviews to gather organisational needs as it provides an opportunity for probing to get detailed information compared to questionnaires (Kajornboon, 2005). We considered the three main different types of interviews: (i) unstructured interviews, (ii) semi-structured interviews and (iii) structured interviews. In this study, we used semi-structured interviews as it is flexible and helps to delve deep into issues. The major objective of this interview is to understand current cyber security training practice in organisations and their needs in terms of the game-based cyber security trainings. The approach for developing the interview guide followed that of the questionnaire. The draft interview guide was reviewed by the authors who were not involved in its design. The interview guide was updated based on the review inputs. The updated interview guide was then applied in the interview of the CISO responsible for choosing cyber security trainings in an organisation (in which we intend to develop a game-based cyber security training). In addition, we interviewed an expert with 19 years of experience in information/cyber security and responsible for choosing cyber security trainings in another organisation in a different country. This is to get a global view on organisational needs by comparing the results from both the interviews. The interview guide is not provided as a part of this paper due to page limit. However, the essence of questions asked in the interviews are provided in Section 4.2.

The results from the survey were extracted and analysed by one of the authors. This was then validated by another author. We tabulated and aggregated the gathered data and then visualised it mainly using graphs and tables as it can be directly translated into a set of high-level requirements on preferences of end-users in a specific organisation. A particular focus was on the preferred training methods and game features. We used content analysis specifically condensation (Erlingsson and Brysiewicz, 2017) as a method to analyse organisational needs and extracted the set of high-level requirements on different game features like ideal playing time, customisability by manually analysing the interview notes.

#### 4. Results: Survey on end-user preferences and interviews on organisational needs

##### 4.1 Survey results: End-user preferences

In this section, we summarise the results from the survey which we conducted as a part of this study. We received 110 responses in total from employees in an organisation to which we sent the questionnaire with 19 questions as shown in Appendix A. No prerequisites were given and not any specific group of employees were asked to answer. Figure 1 shows that 82 out of 110 respondents considered themselves to be at “Newbie/Beginner” level in cyber security. Furthermore, this figure also shows the cyber security level of respondents in two different age categories (i.e., less and more than 40).

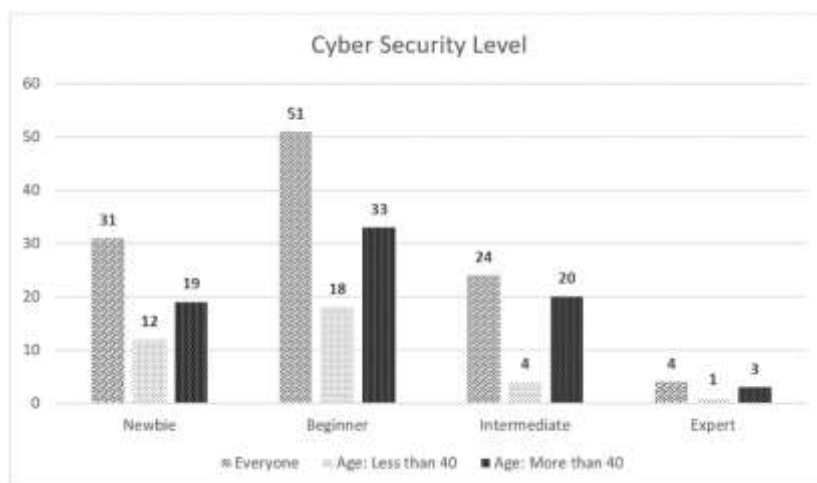


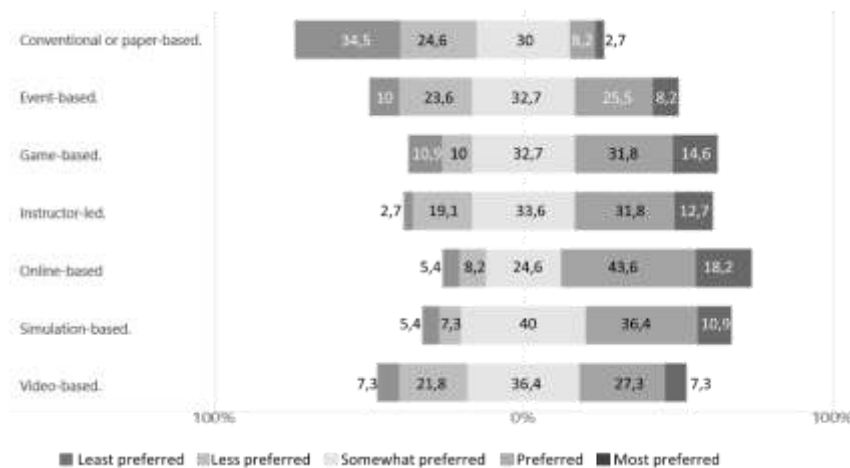
Figure 1: Current cyber security level from respondents perspective

Figure 2 shows that online-based and conventional were the top two training methods used in cyber security trainings that respondents underwent previously. Furthermore, Figure 2 also shows that 33 employees did not undergo any cyber security trainings.



**Figure 2:** Methods used in cyber security trainings that respondents underwent

The respondents were asked to rank each training method from “Least preferred training method” to “Most preferred training method” (Appendix A, Q.11). To avoid any confusion and ensure a common understanding, examples were provided for each category of training method. Figure 3 shows that the conventional (or paper-based) and event-based training methods mostly leaned towards least/less preferred training method. On the other hand, the online-based, simulation-based and game-based training methods mostly leaned towards preferred/most preferred training method. The responses to the question on which factors might lead to a training method success (Appendix A, Q.12) show that, “Active learning process”, “Fun” and “Challenge” were the top three success factors of a training method. On the question of which aspects makes a good game (Appendix A, Q.14): “Combining fun and realism”, “Continuous challenge”, “Interesting storylines”, “Immediate feedback, useful rewards” were highly ranked among the listed aspects of a good game. Based on the responses to question on their preferences to play digitally or non-digitally (Appendix A, Q.15): 72 out of 110 preferred to play digitally whereas 38 out of 110 preferred to play non-digitally. In addition, for question on their preferences to play alone or with others (Appendix A, Q.16): 43 out of 110 preferred to play alone whereas 67 out of 110 preferred to play with others.



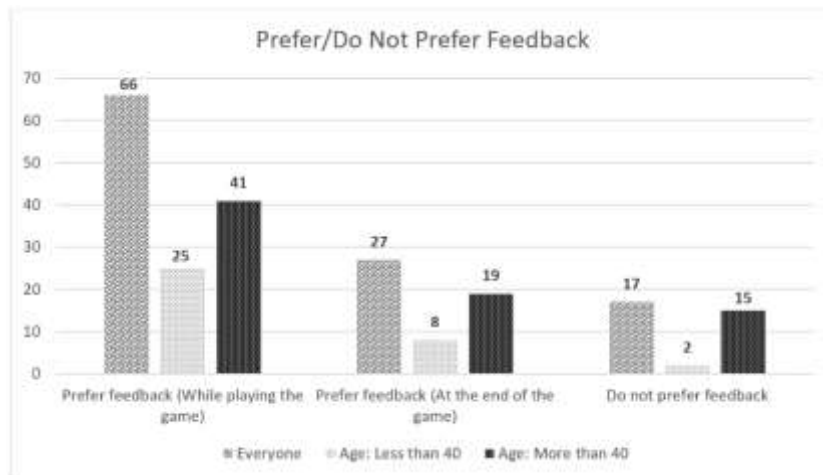
**Figure 3:** Training method preferences

Table 1 shows that 40.9% preferred to play “Team vs. Team” and 27% preferred playing “Co-operative”. On the other hand, only 14.5% liked to play alone against others (i.e., “Player vs. Player”). This indicates that mostly people like to interact with others in games.

**Table 1:** Preferred mode of playing with others

	Player vs. Player	Team vs. Team	Co-operative (Co-op)	Not Applicable	Others
Everyone	14.5%	40.9%	27.3%	15.5%	1.8%
Age: Less than 40	11.8%	38.2%	47.1%	0%	2.9%
Age: More than 40	16.0%	41.3%	18.7%	22.7%	1.3%

Figure 4 shows that 93 out of 110 preferred feedback either during the game or at the end of the game whereas only 17 out of 110 did not prefer feedback. Furthermore, this Figure also specifically shows the preferences in two age categories (i.e., less and more than 40).



**Figure 4:** Respondents preferences on feedback

This analysis has now resulted in a set of high-level requirements: (i) there is a clear need for cyber security training, (ii) the conventional training method is the least preferred training method, whereas the game-based training method is one of the top most preferred/preferred training methods, (iii) the training method should mainly possess factors like “active learning process”, “fun” and “challenge”, (iv) the game should combine fun and realism, give continuous challenge and provide immediate feedback, useful rewards, (v) the game should be mainly playable in a digital platform, (vi) the game should facilitate to interact with others using “Team vs. Team” and “Co-operative” mode of play and (vii) the game should provide feedback (both during the game and at the end of the game).

#### 4.2 Interview results - organisational needs

In this Section, we summarise results of both the interviews performed to elicit organisational needs. A set of requirements that will act as a basis to develop a serious game and fulfil needs of a specific organisation are highlighted. We mainly utilised condensation process to extract key needs. Condensation process helps to shorten the interview text while retaining the core meaning (Erlingsson and Brysiewicz, 2017). After each interview, the interview transcript was reviewed and approved by the corresponding interviewee. The results of the two interviews are categorised on different themes like target group, ideal playing duration as shown in Table 2. The details of the interviewees are not provided due to privacy issues.

**Table 2:** Interview results: Organisational needs

Theme	Interviewee I	Interviewee II
Target Group	(i) Everyone, (ii) Different groups like nuclear, Research and Development (R&D).	(i) Everyone, (ii) Different groups like information technology, R&D.
Customisation	(i) Everyone – general training; Different groups – specialised training, (ii) Elder group – less advanced game features; Younger group – more advanced game features.	(i) Everyone – general training; Different groups – Specialised training.
Topics of Interests	(i) General training – a. how to be aware of and handle phishing emails? b. what to do in terms of cyber security when an employee is travelling? c. what to do if an employee sees something suspicious with regard to their information	(i) General training – general cyber security best practices, (ii) Specialised training – a. Topics on technical aspects of cyber security like how to protect a windows machine, cloud

Theme	Interviewee I	Interviewee II
	assets? (ii) Specialised training – “sikkerhetsloven” is an important topic for the nuclear group in the context of Norway.	infrastructure is important for the IT group.
<b>Training Frequency</b>	(i) General training – monthly once, (ii) Specialised training – “need-to-be-done” basis (legislations, customer agreements).	Monthly once.
<b>Ideal Playing Duration</b>	30 minutes (Divide the game into modules which would allow them to do it step-by-step).	(i) 10 – 30 minutes (non-digital game), (ii) 5 – 15 minutes (digital game: each module).
<b>Trainee Evaluation</b>	(i) General training – checklist, (ii) Specialised training – exam to evaluate whether the training content improved their knowledge and understanding.	Use of progress indicators (by checking actual pace against the right pace of the awareness action plan) and efficiency indicators (through quiz, simulation).
<b>Training Evaluation</b>	Required.	Required – use of feedback forms.
<b>Basis for Training Update</b>	Required – use of threat/risk picture of our organisation, incidents around the world, feedback from trainees.	Required – recent indicators like increasing type of cyber-attacks during Covid-19, data changes.

The following organisational needs were highlighted by both interviewees, which might reflect a global view: (i) cyber security trainings in an organisation should be tailored to at least two different groups i.e., for everyone and for specific teams like R&D, (ii) cyber security trainings need to be conducted at least once every month, (iii) trainees should be evaluated once they complete a cyber security training, (iv) cyber security trainings (including training content) should be evaluated by the trainee using feedback forms as it could be used as a basis for training updates and (v) cyber security trainings (including game content) needs to be updated periodically by using the risk picture corresponding to that organisation and also considering the global trend on different cyber-attacks. Furthermore, the interviewee 1 who is a part of the organisation in which we intend to develop a serious game for cyber security training, highlighted some additional needs specific to their organisation which include: (i) game features should be customisable for different age groups (younger group – more advanced, elder group – less advanced), (ii) ideal playing duration should be 30 minutes and (iii) the game should be composed of different connected modules with 30 minutes for each module.

## 5. Discussion

This section discusses implications, generalisability and limitations of this study.

In the survey which we conducted, most of the respondents reported that they underwent training using conventional method. In contrast, this was also the method with the least preference. This implies that respondents prefer other training methods compared to conventional method for cyber security trainings. For some of the questions, data indicated a clear difference between those above 40 years of age and those under, concerning both their self-evaluation and their preferences related to game features. For instance, regarding cyber security level, 26.6% of respondents in the age group above 40 considered themselves as “intermediate” versus 11.4% of respondents in the age group less than 40. On the other hand, 51.4% of respondents in the age group less than 40 considered themselves as “beginners” versus 44% respondents in the age group above 40. Similarly, there is also a difference in preferences to play alone or with others (either as a team or against each other). Results of the survey also shows that the game-based method and online interactions were more preferred by respondents in the age group less than 40. This advocates the use of game-based methods in trainings. A central question here is: “Whether to cater to everyone’s needs or focus on the preferences of the major group as it seems this will be the prevalent group in the years to come?”. In terms of gaming features, most of the respondents preferred active learning with challenges, fun and feedback. The game-based method seems to be one of the promising methods to use in cyber security trainings which was also echoed by the interviewees. Furthermore, most of the respondents preferred to play in teams and interact with others. This also supports team building exercises in organisations. Finally, an option to play digitally is mostly preferred which is also appropriate considering the Covid-19 pandemic.

The survey was applied to one specific organisation in Norway and the results would not directly represent employees in different organisations in Norway or in other countries. However, the respondents had a good mix in terms of gender, age, and sector they work in or study towards. Furthermore, most of the respondents had a University degree, which also seems to be typical in other organisations. In addition, we did not perform any

tests of statistical significance as the major goal of the survey was to determine the subjective needs of the potential end-users and specific organisation. However, we presented some of the results to the CISO whom we interviewed. The general reflection provided was that self-evaluation of cyber security level is as expected for their organisation. Furthermore, “playing with others” seems to be an appropriate choice for their organisation. Finally, it is also important for their organisation to find a good balance between training/game and day-to-day business activities. This also makes it appropriate to consider organisational needs on features like ideal playing duration. Moreover, results of the survey also showed correspondence with literature. For instance, the top three success factors of a training method (“Active learning process”, “Fun” and “Challenge”) closely reflects the findings of (Ghazvini and Shukur, 2017) on training success factors. In addition, the method which we used to elicit and analyse end-user preferences and organisational needs can be directly used/adapted in different organisations in Norway or organisations in other countries. This indicates the generalisability on different aspects of this study. Even though we did not focus on specific groups in an organisation, this survey covered end-user preferences on training methods and game features which is the key for developing a serious game. The training content can be adapted later for general and specialised training. Typically, there are a limited number of people in the management responsible for choosing cyber security trainings in an organisation. In addition, there is a limited time availability, which resulted in one interview with the CISO who is responsible for choosing the cyber security trainings in an organisation. However, we complemented it with another interview of an expert in another organisation in a different country to ensure the quality and get a global view on different aspects.

Due to the Covid-19 pandemic, we relied only on using questionnaire for eliciting end-user preferences. The respondents were aware that their responses would have an impact on the management choice in the type of training methods and game features to be used in the subsequent cyber security trainings. This in turn motivated the respondents to take this questionnaire seriously. Furthermore, we mainly utilised closed-ended questions in the questionnaire which also helps to prevent respondents from providing invalid data. However, in the future, we could utilise multi-methodology approach for data collection. This implies that, in addition to questionnaire, we could also use interviews and/or focus groups. This in turn will help to gather more comprehensive data.

## **6. Conclusions and future work**

Existing serious games in cyber security follows one-size-fits-all approach which results in unwanted game features and lack of relevant storylines that cater specific and varying needs of an organisation. This in turn lead to ineffective cyber security trainings in practice. Therefore, in this paper, we proposed an approach to elicit end-user preferences using surveys and needs of a specific organisation through interviews. Based on this approach, we gathered preferences of 110 potential end-users in a specific organisation. The analysis of the gathered data led to a set of high-level requirements on different game features like prefer/do not prefer feedback, prefer digital/non-digital game. On the other hand, we interviewed key personnel to mainly gather needs of a specific organisation. The analysis of interview transcript using the condensation process led to a set of high-level requirements on different organisation specific needs like topics of interest, ideal playing duration. In the future, the set of high-level requirements gathered as a part of this study would act as a basis to develop a serious game for cyber security training in a specific organisation. Furthermore, we intend to perform a comparative analysis on the effectiveness of an existing serious game in cyber security with a serious game developed based on the end-user preferences and organisational needs.

## **Appendix A. Cyber Security Training: Experiences and Preferences of End-users**

1. What is your gender?

- Female
- Male
- Not listed above
- Prefer not to say

2. What is your age group?

- Under 18
- 18 - 24

25 - 39

Above 40

3. What is the highest level of education you have completed?

Secondary School

High School

University Degree (Bachelor's/Master's/PhD)

Others, please specify:

4. Which sector do you work in or study towards? Please select all that apply.

Education

Energy

Financial

Healthcare

Process Industry

Public Sector

Transport

Research

Information Technology (IT)

Others, please specify:

5. According to you, what is cyber security?

6. Why do you think organizations need cyber security?

7. Who is responsible for cyber security in your organization?

Chief Executive Officer (CEO)

Chief Security Officer

Line Manager

Chief Information Security Officer (CISO)

ICT Department

All Employees

All of the Above

8. I know my department's digital assets, their value for our business, their vulnerabilities, their threat picture and how to protect them against cyber threats.

Strongly

Disagree

Neutral

Agree

Strongly

9. What is your cyber security level?

Newbie: I have no experience and not completed any course in cyber security

Beginner: I have little experience and/or completed a basic course in cyber security

Intermediate: I have some experience in cyber security and/or completed several courses in cyber security

Expert: I have considerable experience in cyber security and/or teach cyber security on a professional level

10. Which of the following methods were used in the cyber security training you underwent? Please select all that apply.

Conventional (or paper-based)

Event-based

Game-based

- Instructor-led (or classroom-based)
- Online-based
- Simulation-based
- Video-based
- None of the above
- Others, please specify:

11. What is your preference for each of the following training methods?

	Least preferred training method	Less preferred training method	Somewhat preferred training method	Preferred training method	Most preferred training method
Conventional or paper-based.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Event-based.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Game-based.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instructor-led.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online-based.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Simulation-based.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Video-based.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

12. Which of the following do you think are the success factors of a training method? Please select all that apply.

- Passive learning process
- Active learning process
- Challenge
- Easily accessible
- Fun
- Motivation
- Multiple topic coverage
- Others, please specify:

13. Which type of games do you like? Please select all that apply.

- Adventure games
- Role-playing games
- Simulation games
- Strategy games
- Others, please specify:

14. Which of the following features do you think makes a good game? Please select all that apply.

- Combining fun and realism
- Continuous challenge
- Flexibility
- Interesting storyline
- Immediate feedback, useful rewards
- Others, please specify:

15. What is your most preferred platform to play games?

- Non-digital
- Digital

16. Do you like to play games alone or with others?

- Alone
- With others

17. In case you would like to play games with others, what is your most preferred option?

- Player vs. Player



- Team vs. Team
- Co-Operative (Co-Op)
- Not applicable
- Others, please

18. Do you like to have feedback to monitor your progress in the game?

- Yes, I would like to have feedback while playing the game
- Yes, I would like to have feedback at the end of the game
- No, I would want to play the game for fun and learn without any feedback

19. Considering the definition of a "Serious Game" as "games whose primary objective is not fun or entertainment, rather learning or practicing a skill". What is your experience with playing serious games?

- I never played serious games and have no experience
- I played serious games once or twice and have some experience
- I played serious games several times and I am quite experienced
- I played serious games extensively and I am an expert

## References

- Abawajy, J. 2014. User Preference of Cyber Security Awareness Delivery Methods. *Behaviour Information Technology*, 33, 237-248.
- Alotaibi, F., Furnell, S., Stengel, I. & Papadaki, M. 2016. A Review of using Gaming Technology for Cyber-security Awareness. *Int. J. Inf. Secur. Res.*, 6, 660-666.
- Ávila-Pesantex, D., Rivera, L. A. & Alban, M. S. 2017. Approaches for Serious Game Design: A Systematic Literature Review. *The ASEE Computers in Education (CoED) Journal*, 8.
- BBC-News. 2020. Norway Blames Russia for Cyber-attack on Parliament [Online]. Available: <https://www.bbc.com/news/world-europe-54518106> [Accessed 03.12.2020].
- Chittaro, L. & Ranon, R. Serious Games for Training Occupants of a Building in Personal Fire Safety Skills. 2009 Conference in Games and Virtual Worlds for Serious Applications, 2009. IEEE, 76-83.
- Erlingsson, C. & Brysiewicz, P. 2017. A Hands-on Guide to Doing Content Analysis. *African Journal of Emergency Medicine*, 7, 93-99.
- Ghazvini, A. & Shukur, Z. 2017. A Framework for an Effective Information Security Awareness Program in Healthcare. *International Journal of Advanced Computer Science Applications*, 8, 193-205.
- Graafland, M., Schraagen, J. M. & Schijven, M. P. 2012. Systematic Review of Serious Games for Medical Education and Surgical Skills Training. *British Journal of Surgery*, 99, 1322-1330.
- Kajornboon, A. B. 2005. Using Interviews as Research Instruments. *E-journal for Research Teachers*, 2, 1-9.
- Lallie, H. S. et al., 2020. Cyber Security in the Age of Covid-19: A Timeline and Analysis of Cyber-crime and Cyber-attacks during the Pandemic. arXiv preprint arXiv:2006.11929.
- Lee, R. M., Assante, M. J. & CONWAY, T. 2014. German Steel Mill Cyber Attack. *Industrial Control Systems*, 30, 62.
- Offermann, P., Levina, O., Schonherr, M. & Bub, U. Outline of a Design Science Research Process. *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, 2009. ACM, 7.
- Pfleeger, S. L., Sasse, M. A. & Furnham, A. 2014. From Weakest Link to Security Hero: Transforming Staff Security Behavior. *Journal of Homeland Security and Emergency Management*, 11, 489-510.
- Ponsard, C. & Grandclaoudon, J. Survey and Guidelines for the Design and Deployment of a Cyber Security Label for SMEs. *International Conference on Information Systems Security and Privacy*, 2018. Springer, 240-260.
- Sfakianakis, A. et al., 2019. ENISA Threat Landscape Report 2018: 15 Top Cyberthreats and Trends. 10.
- Shostack, A. 2018. Tabletop Security Games & Cards [Online]. Available: <https://adam.shostack.org/games.html>.
- Thomas, J. 2018. Individual Cyber Security: Empowering Employees to Resist Spear Phishing to Prevent Identity Theft and Ransomware Attacks. *International Journal of Business Management*, 12, 1-23.
- Tioh, J.-N., Mina, M. & Jacobson, D. W. Cyber Security Training A Survey of Serious Games in Cyber Security. 2017 IEEE Frontiers in Education Conference (FIE), 2017. IEEE, 1-5.
- Van Ruijven, T. Serious Games as Experiments for Emergency Management Research: A Review. *ISCRAM 2011: Proceedings of the 8th International Conference on Information Systems for Crisis Response and Management*, Lisbon, Portugal, 8-11 May 2011, 2011. ISCRAM.
- Verizon 2019. Data Breach Investigations Report.
- Wattanasoontorn, V., Boada, I., Garcia, R. & Sbert, M. 2013. Serious Games for Health. *Entertainment Computing*, 4, 231-247.

# Effectiveness of Covert Communication Channel Mitigation Across the OSI Model

Tristan Creek, Mark Reith and Barry Mullins

Air Force Institute of Technology, Wright-Patterson AFB, USA

[Tristan.Creek@afit.edu](mailto:Tristan.Creek@afit.edu)

[Mark.Reith.ctr@afit.edu](mailto:Mark.Reith.ctr@afit.edu)

[Barry.Mullins@afit.edu](mailto:Barry.Mullins@afit.edu)

DOI: 10.34190/EWS.21.108

**Abstract:** The Internet consists of various levels of communication technologies which are often categorized by a layered model called the OSI model. Among the technologies within the seven layers of the OSI model, covert communication channels allow attackers to subvert defenders and secretly transmit data by leveraging technologies beyond their specified standards. These covert communication channels impact security differently depending on which layer of the OSI model they exist. Although mitigating every channel is ideal, limited resources require the consideration of the effectiveness of mitigating covert communication channels. This paper presents the impact of covert communication channels on each layer of the OSI model, how to mitigate them, and concludes with recommending the mitigation of covert communication channels on layers 1 through 4 of the OSI model. The final recommendation deliberates that an organization must decide their cost-to-risk ratio on an individual basis when considering solutions to mitigating covert communication channels.

**Keywords:** covert communication channels, OSI model, network technology

## 1. Introduction

Interconnected computer systems share information to enable all functions of the Internet from articles curated in Facebook feeds to industrial control of water towers that maintain healthy water pressure for entire cities. One would be correct to assume that these two seemingly disconnected functions utilize different technologies to achieve their goals, but many may be unaware of their underlying similarities. Despite their apparent surface-level differences, internet connected systems of great variety utilize similar communication technologies that can be better understood with a layered model.

A layered model provides the benefit of modelling network communication systems from the top down, with the lowest layers making up the most basic, shared technologies. The higher layers represent the most specific technologies, such as rendering an article in a web browser or interpreting data to control a water tower pressure pump, and build upon the lower layers. This concept of layers is more formally presented in the Open System Interconnection (OSI) model which describes seven different layers of network communication technologies as seen in Figure 1. (Li, Li, Cui, & Rui, 2011).

7	Application Layer	Human-computer interaction layer, where applications can access the network services
6	Presentation Layer	Ensures that data is in a usable format and is where data encryption occurs
5	Session Layer	Maintains connections and is responsible for controlling ports and sessions
4	Transport Layer	Transmits data using transmission protocols including TCP and UDP
3	Network Layer	Decides which physical path the data will take
2	Data Link Layer	Defines the format of data on the network
1	Physical Layer	Transmits raw bit stream over the physical medium

**Figure 1:** The seven layers of the OSI model categorize various communication technologies (Imperva, 2020)

The varying technologies categorized in each of the seven layers utilize standardized specifications to ensure all systems can communicate with each other if they implement the technology and use it based on the specifications. These specifications are also of critical importance to network defense because the format of data must be understood in order to interpret and monitor data at any given layer. Even if defense tools make use of these specifications, attacker may exploit the specifications to covertly communicate data in unintended ways to subvert defenses.

A covert channel (referred to in this paper as ‘covert communication channel’) is “any communication channel that can be exploited by a process to transfer information in a manner that violates the system's security policy” (Defense, 1985). As this paper presents, covert communication channels exist on every layer of the OSI model and present serious concern for security. Securing communication technologies to mitigate all of these covert communication channels is ideal, but finite resources may not allow for this. Therefore, this paper seeks to analyze the general trends of covert communication channels across the seven layers of the OSI model to identify at which layers it would be best to mitigate covert communication channels for maximum effect. To accomplish this, the following sections briefly introduce each layer, provide an example of a covert communication channel per layer, discuss mitigation, and conclude with an overall analysis and mitigation recommendations.

### 1.1 Layer 1 – physical layer

The lowest layer of the OSI model, the physical layer, encapsulates the rawest form of data transfer between systems. The physical layer includes the hardware required to send and receive network data, most notably the network interface card (NIC). These hardware devices are responsible for transmitting data to other systems across a physical communication channel such as an ethernet cable or through the air for technologies like Bluetooth (Li, Li, Cui, & Rui, 2011).

On the topic of Bluetooth, this wireless communication protocol allows devices to communicate with each other over the air rather than through a wire like traditional computers. Although certain Bluetooth protocols require line-of-sight for direct transmission between two devices, researchers developed methods to reflect or diffuse the Bluetooth signal off of surfaces to bypass this line-of-sight requirement as seen in Figure 2 (Liu, Liu, Zeng, & Ma, 2020). This ability provides complete control of informational direction that prevents a third-party from interfering with the signal, meaning a defending party cannot intercept the signal.

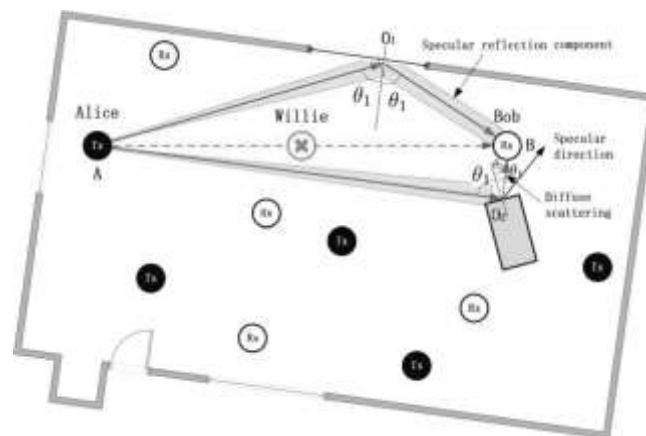


Figure 2: Reflecting or diffusing Bluetooth signal to avoid third-party interdiction (Liu, Liu, Zeng, & Ma, 2020)

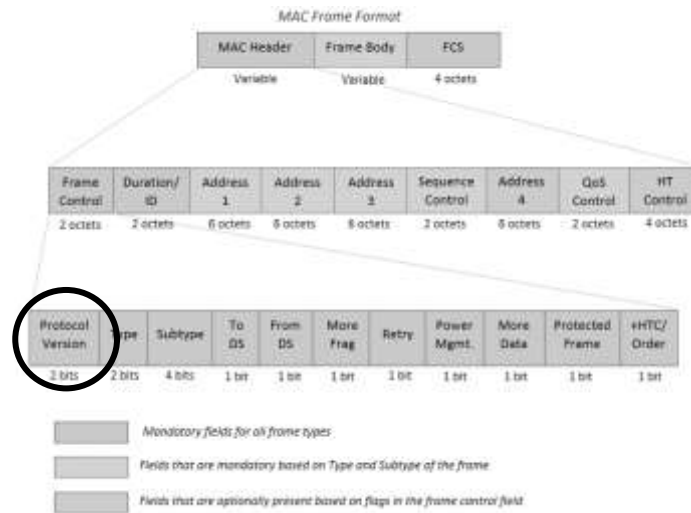
This fundamental control of the physical medium allows attackers to setup a covert communication channel. This would prove a more serious risk if it allowed attackers to subvert interception on international physical mediums such as undersea fiber cables, but the short range of the attack coupled with the difficulty of preventing it places this covert communication in a low priority for mitigation. The layer above the physical medium is called the Data Link Layer as follows.

### 1.2 Layer 2 – data link layer

The second lowest layer of the OSI model, the data link layer, is responsible for encapsulating outgoing data and decapsulating incoming data based on media access control (MAC) frame format standards. The MAC standard includes sender and recipient data in the header as well as meta data for protocol version identification and

error checking (Li, Li, Cui, & Rui, 2011). The sender exercises full control of the meta data which can be used to transmit data in unexpected ways.

The Protocol Version field specified by the MAC frame standards consists of only two bits as seen in Figure 3. The standards specify that only the value 0 is ever used in practice because 1, 2, and 3 are reserved (IEEE Standard for Ethernet, 2018). Therefore, this field does not contain critical data since the value can be assumed to be 0, leaving this field open for custom data storage for a covert communication channel. (Gonçalves, 2011).



**Figure 3:** MAC header with protocol version in bottom left (MathWorks, 2020)

Unlike the raw data transmitted through a physical media on layer 1, layer 2 protocols provide more specific standards that allow defenses to check the validity of traffic and prevent covert communication channels. In regards to the provided example, a network defense tool simply must ensure that the Protocol Version field on each MAC header equals 0 and reject otherwise. This generalizes the covert communication channel attacks and defenses on this layer by some bit, but the ability to ensure protocols meet the defined standards provides strict defensive measures across layer 2.

### 1.3 Layer 3 – network layer

The third layer of the OSI model, the network layer, serves the function of packet transmission across hosts. The network layer most commonly transmits internet protocol (IP) packets which contain source and destination IP addresses for routing purposes. Although the second layer facilitates communication between two devices, this network layer handles navigation across multi-hop networks so systems can reach far away destinations (Li, Li, Cui, & Rui, 2011).

Other researchers have developed a proof-of-concept tool called Covert\_TCP to demonstrate a covert communication channel implemented in the identification (ID) field of an IP header as seen in Figure 4. This 16-bit ID field can store two characters or, in this proof of concept, one encoded character for added stealth. This is a basic example that stores data in a header field similar to the data link layer covert communication channel, but more advanced, stealthy channels are possible on the network layer (Sbrusch, 2006).

The Internet Protocol uses the ID field to recombine data that the sender split into chunks, but unsplit data may set the ID field to an arbitrary value (Touch, 2013). Therefore, mitigating covert communication channels on this layer requires more work than just validating that a protocol follows the standards. In this case, behavioral analysis may be employed to detect suspicious behavior such as purposefully sending many small pieces of data in order to control the ID field rather than sending one large piece of data that splits and relies on the ID field. Although more difficult than the simple protocol standard matching from layer 2, the ability to spot behavioral outliers is a necessity among more complex technologies with complex specifications.

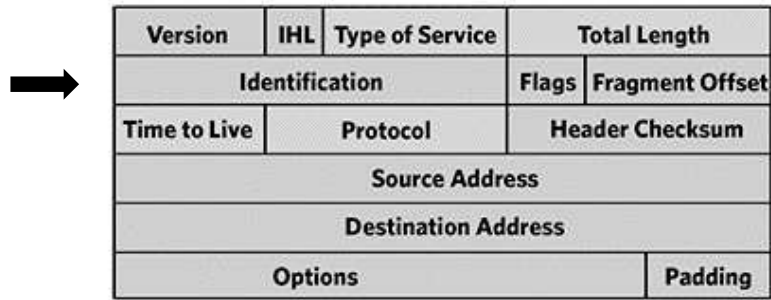


Figure 4: IP header with identification field at the beginning of the second row (Salutari, 2020)

### 1.4 Layer 4 – transport layer

The fourth layer of the OSI model, the transport layer, provides many useful functionalities on top of the network layer. These functionalities include packet ordering, re-transmission, error identification and correction, and more. The most popular protocol in this layer, the Transmission Control Protocol (TCP), builds on top of the IP protocol from the network layer. Often referred to as TCP/IP, this layer 4 protocol includes exploitable header fields similar to IP (Li, Li, Cui, & Rui, 2011).

The aforementioned proof-of-concept tool Covert\_TCP also provides functions for covert communication channels on the transport layer via TCP header fields. Rather than the ID field of the IP header, however, the tool stores a character in the Sequence Number (SN) field of the TCP header as seen in Figure 5. Senders and receivers use this field to identify packet loss and correct order of packets, but a sender can avoid dependence on this field. By never fully establishing a connection, a sender can continuously send packets with full control of the data in the SN field (Sbrusch, 2006).

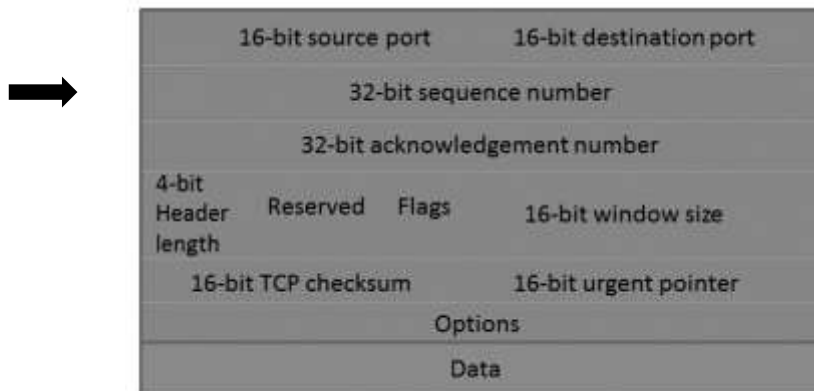


Figure 5: A TCP header contains a sequence number field on the second row above (ComputerNetworkingNotes, 2020)

The mitigation of this covert communication channel falls directly in line with the mitigation of the layer 3 covert communication channel. Despite this example following the TCP standard, defenders can mitigate this communication method by recognizing the uncharacteristic amount of TCP packets attempting to start a connection without every establishing a connection. This presents a method for behavioral detection that leverages common trends among systems and users to identify outliers and reject them.

### 1.5 Layer 5 – session layer

The fifth layer of the OSI model, the session layer, manages the active session between two connected clients. Just like the transport layer builds upon the network layer, this session layer builds upon the transport layer and provides flow management options. (Li, Li, Cui, & Rui, 2011).

The session layer Real-Time Transport Protocol (RTP) has a sister protocol called the RTP Control Protocol (RTCP) which provides control over an RTP session. RTCP packets contain an interarrival jitter field to store an estimate of the time between packet arrivals to manage the session flow. Similar to the covert communication channels

mentioned for previous layers, simply storing a character in the interarrival jitter field, as seen in Figure 6, works but lacks stealth. However, an attacker can analyze the natural jitter and encode a character to produce a value that looks similar to a natural jitter value. (Bai, Huang, Hou, & Xiao, 2008).

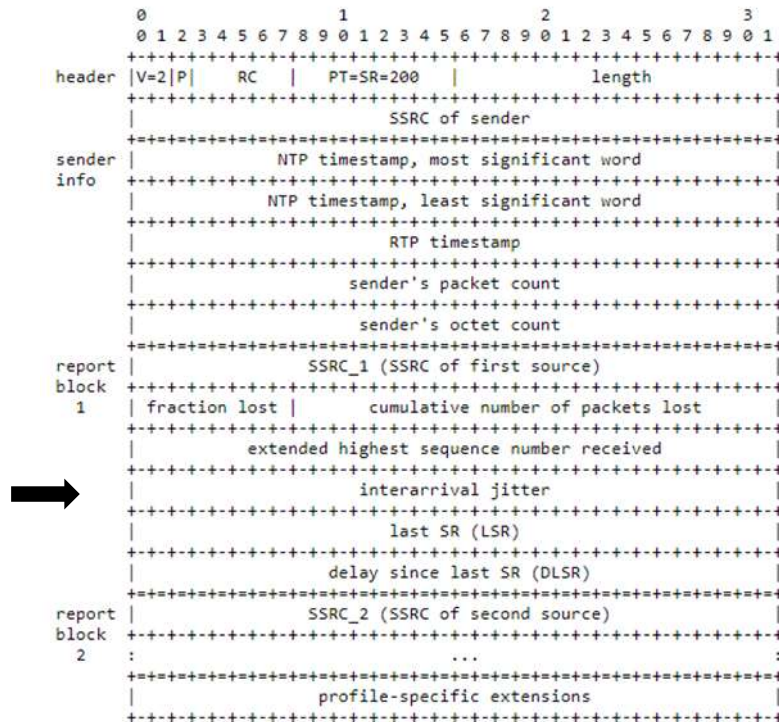


Figure 6: RTCP packet format with the jitter field labeled "interarrival jitter" (Schulzrinne, Casner, Frederick, & Jacobson, 2020)

In this example, behavioral analysis fails to identify outlying jitter values when the sender encodes them to look natural. This presents challenges beyond just ensuring protocols match their standards and identifying behavioral outliers, now requiring defenders to internally control technologies for lack of ability to identify suspicious data in transit. Therefore, defense becomes more difficult as it must address individual system security rather than the detection of covert communication channels on centralized network systems through which data travels.

### 1.6 Layer 6 – presentation layer

The sixth layer of the OSI model, the presentation layer, handles the presentation of data and changes to be made to it. These changes include encrypting/decrypting and format conversion such as binary to text. This layer reaches closer to the user than the lower layers do, and again introduces more complexity with the availability of more technologies (Li, Li, Cui, & Rui, 2011).

In regards to the presentation of data, a covert communication channel called image steganography describes the general process of hiding data within images in a manner unnoticeable to the naked eye. A popular method of image steganography is least significant bit (LSB) manipulation which stores data in the LSB of the three color components of a pixel, as seen in Figure 7. In practice, storing data in the LSB of each pixel changes the color so little that the new image appears the same as the original image to the naked eye (Poornima & Iswarya, 2013).

Due to the variety of ways an attacker could encode the data they wish to covertly communicate over this channel, attempts to rebuild the data encoded in an image would require dedicated resources and likely identify few covert communications even if they existed in the image. This presents challenges similar to those identified in the covert communication channel on layer 5; defenses on this layer should extend to individual systems to prevent the storage and transmission of sensitive data within images since identifying this covert communication at central network points proves too difficult.



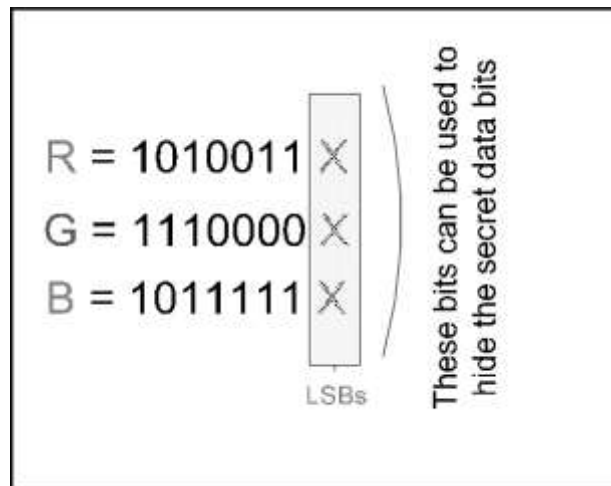


Figure 7: The least significant bits of the three color components of a pixel (Altigani, Hasan, & Barry, 2020)

### 1.7 Layer 7 – application layer

The seventh layer of the OSI model, the application layer, is responsible for interfacing with applications on a system to allow them to utilize network services and communicate with other systems that use the same protocol. This layer includes popular user-facing protocols such as the Hypertext Transfer Protocol (HTTP) which serves website pages to browsers (Li, Li, Cui, & Rui, 2011). In relation to HTTP, an interesting covert communication channels arises under the Domain Name System (DNS) protocol which resolves website domains to IP addresses so users can connect to them. Botnet owners have been identified using deterministic domain generation algorithms (DGAs) to facilitate command and control of their bots. Not only are the domains generated by a DGA difficult to predict and block due to their random nature as seen in Figure 8, but botnet owners also often hide the DNS requests to these domains by encrypting the traffic with the HTTP secure (HTTPS) protocol. (Patsakis, Casino, & Katos, 2019). This covert communication channel allows botnet owners to subvert detection and misuse DNS to control infected systems.

```
"mct2v81ktg4211kq03sy1rm0uxo.net",  
"1bouh8d1qq1gw1sovuxy1vet1pz.com",  
"booxhk1k2uvbhj77q5ypcyoaj.net",  
"5egzd415k5my9ejjuju1dqnhzt.org",  
"o4882k1dc0h3js8xw231rzu6ie.org",  
"ox60c0gnucrefm6zz11cnk3q8.com",  
"16kxx7t1bz4jgm11jeulq1ewe58s.org",  
"1hm0718oqkd2w16o6fb7akrmg.com",  
"152dggc1ut6ez59emnwalc26on.com",  
"1fyg4aa1un495ctjmp1xlti6d.net",  
"xxi1gggdio4i1dwbqihupm0uk.com",  
"194h5uc1k2y1qh15rfwlib78r74.net",  
"18d3eqg1f99immh3lyge1uz4cvq.biz",  
"13k9kj11ucaywg1gacucw8q6sc4.net",  
"18qawq1v9cgty12nbbvxssouim.net",  
"152pz101i1z4wza1rg9f1xq0ymh.com",  
"5utd8gb60eaw3espa91lhkz64.org",  
"1v4ch091mdfeov1qi9iid49o9ow.net",  
"n03kh4jmrjjhdycsv4a17g2e.com",  
"15unmxej55yrcvdknsj1reo6g2.net",  
"1dxq591102z0gkxzkrw4d59zw8.com",  
"nr0doipw73cmt5qvz01okg6bo.net",  
"npxarpozgy2t683db11992r4d.net",  
"1bxuqto1op5yyc1e7c4pt1gqztu1.org",  
"s0zrt11nfsge7yorbwa71401y.com",  
"1ki2qg6z4pvbbuoonp2b6g0j.biz",  
"4qgugc1pixqzxr1ka0ov8yac0.biz",  
"hn0r1ltqiqvzsf1r6y49bf8.com",  
"1pxs0y9ttt0g1sqe4wgypuwkq.biz",
```

Figure 8: Domains generated by a domain generation algorithm (Ahuje, 2020)

Encrypting traffic with HTTPS hides the DNS requests on typical networks, but internal enterprise networks can decrypt traffic to analyze before re-encrypting and forwarding it (O'Neill, M. et al. 2017). Similar to layers 5 and 6, this requires involving individual systems in the defense solution.

## **2. Mitigation analysis**

As presented in the final paragraph of each OSI layer section, mitigation of covert communication channels can be categorized as follows:

### **2.1 Layer 1 - physical layer**

If an attacker possesses control of, or access to, the physical medium to enable covert communication channels, defenders have little chance to mitigate the threat. Therefore, mitigation of covert communication at this layer requires controlling access to the physical mediums (Bluetooth signals, ethernet cables, switches, etc.) and preventing access to unauthorized users. In order to maintain control of wireless signals such as the Bluetooth example presented in Section 1.1, radiofrequency (RF) protective barriers exist to separate protected internal signals from external signals or listeners. These barriers are shown to effectively block popular signal ranges such as Bluetooth, WiFi, and Global Positioning System (GPS) (Yadav, Jain, & Sharma, 2019). Some RF barriers also block broad ranges of the RF spectrum as seen in barriers applied in secure compartmentalized intelligence facilities (SCIFs) (Cofer, 2019).

SCIFs adhere to the National Security Agency (NSA) framework codenamed TEMPEST which leverages RF barriers, among many other things, to block unintended electromagnetic signal transmission. TEMPEST provides effective measures to prevent covert communication at the physical layer (Goodman, 2021). The extensive measures put forth by these requirements should be thought of as reasonable means to prevent adversarial covert communication at the physical layer to the best of current abilities. Unfortunately, these physical measures come with a hefty price of anywhere from \$400,000 to \$1 million for a 2,000 square foot SCIF (Walsh, 2007). Fortunately, we will see that mitigation measures cost less at higher OSI layers.

### **2.2 Layer 2 - data link layer**

Much research addresses layer 2 spoofing attacks (Bhaiji, 2007; Convery, 2002), but few sources cover covert communication channels, which are attacks in their own right. Detection and prevention of these covert communication channels relies on the compliance to established protocol formats as detailed in the aforementioned MAC header example from Section 1.2. Deviance from protocol specifications indicates abnormal behavior which warrants logging and potentially blocking based on further analysis. Analysis of layer 2 traffic can be performed manually with tools like Ettercap to detect malicious behavior such as address resolution protocol (ARP) poisoning (Majidha Fathima & Santhiyakumari, 2021).

Even better, next generation firewall (NGFW) hardware devices produced by various companies possess the ability to automatically analyze and allow/block layer 2 traffic based on a specified profile (Forcepoint LLC 2018). Unfortunately, it is difficult to profile every single malicious traffic flow ahead of time for fingerprinting later, and relying on whitelists of known good traffic flows is impractical in large networks with diverse types of traffic. This lack of flexibility paired with the total cost of owning a NGFW (\$58,000 - \$500,000+ (Skybakmoen, 2018)) limits the effectiveness of mitigating covert communications at this level. However, we see the move towards cheaper, more accessible software solutions as we move up the layers of the OSI model.

### **2.3 Layer 3 - network layer / Layer 4 - transport layer**

Although the aforementioned NGFWs also detect and prevent malicious traffic at layer 3 and layer 4, less expensive solutions exist. Rather than rely on expensive hardware, software installed on common switches and routers can analyze traffic. Compared to the cost of NGFWs, this software provides a much less expensive solution where in some cases free software like Snort installed on a less than \$100 raspberry pi computer works as a layer 3 and layer 4 intrusion detection system (IDS) and intrusion prevention system (IPS) (Coşar & Karasartova, 2017).

This inexpensive setup is unfortunately complemented by lower effectiveness. Park and Ahn showed that Snort and Suricata (another free IDS/IPS software) only detect 73% of common malicious traffic at best, and only 16% at worst (2017). On the other hand, the majority of tested NGFWs not only detected but blocked 90%+ of



malicious traffic (Skybakmoen & Dhanraj, 2017). This disparity between effectiveness rates relates directly to the disparity of cost between free software solutions and NGFWs which leaves defenders to decide on the level of security they are willing to pay for.

## **2.4 Layer 5 - session layer / Layer 6 - presentation layer**

As we move to higher layers of the OSI model, the available protocols provide malicious actors better ways to evade detection. Despite the high detection and blocking rates of NGFWs, reports show that they often fail to detect and block malicious traffic employing evasion techniques at higher layers of the OSI model. In testing, the Juniper SRX 4200 blocked 99.24% of all attacks, yet failed to identify malicious traffic that fragmented RPC (a layer 5 protocol) data to evade detection (Skybakmoen & Dhanraj, 2017). Furthermore, detecting image steganography as presented in the LSB example in Section 1.6 proves difficult due to the various methods of encoding data in an image such as encoding data in the second LSB instead of the LSB. The wide range of encoding possibilities available to just image steganography alone shows the difficulty of detecting layer 5 and layer 6 evasion techniques in network traffic. Therefore, the focus at these layers shifts from network-based detection and prevention to host-based detection and prevention.

Rather than detecting image steganography on the network, host-based detection and prevention aims to block tools that enable image steganography. Blocking every image steganography proves difficult when steganography Python scripts are readily available on Github and easily found through a simple Google search (Prado, 2017). Host-based solutions cannot reasonably block all software that enables evasion because malware often uses legitimate protocols (seen in the example in Section 1.5) and legitimate applications including antiviruses (Anthony et al. 2012). Therefore, host-based solutions must instead block the malicious actor's access to and control of the host system. This falls outside the scope of this paper, but illustrates the challenges covert communication channels pose at higher OSI layers.

## **2.5 Layer 7 - application layer**

Encrypting traffic with HTTPS hides DNS requests on typical networks (as discussed in Section 1.7), but NGFWs allow private networks to decrypt and analyze traffic before re-encrypting and forwarding it. This technology, known as SSL/TLS inspection (O'Neill, M. et al. 2017), eliminates the issue of encryption hiding covert DNS communication, allowing machine learning algorithms to detect 97% of unencrypted, malicious DNS requests (Saali, S. et al. 2019). However, this detection of one common evasion technique proves little considering that 6 out of 11 NGFWs analyzed by Skybakmoen and Dhanraj failed to protect against malicious traffic utilizing HTTP evasion techniques (2017).

Similar to the challenges posed by covert communication channels at layer 5 and layer 6, the failure of network detection and prevention requires the focus to shift to host-based detection and prevention instead. This, once again, falls outside the scope of this paper. However, it shows the importance of host system control even though the goal is to detect and prevent hidden network communication.

## **3. Recommended mitigations**

Our analysis from Section 2 identifies two high level themes: mitigations for covert communication channels are only possible at lower levels of the OSI model (specifically layers 1-4), and mitigations for covert communication channels cost more at lower levels of the OSI model (specifically layers 1 and 2). With these two themes in mind, an organization would best serve their needs by weighing the sensitivity of their data and cost of an attacker stealing said data against the cost of infrastructure to stop the attackers. A small business working with private, but not necessarily sensitive, data would likely be best served by an inexpensive Snort IDS/IPS installation and an inexpensive host-based antivirus. On the other hand, a nation-state that handles sensitive nuclear information would likely be best served by implementing everything possible including a SCIF, NGFW with network-based IDS/IPS, and a host-based antivirus.

## **4. Conclusion**

The consideration of which mitigations, if any, to implement can only be decided by an organization's cost-to-risk tolerance. Although covert communication channels will likely always be prevalent, these presented levels of mitigation at least challenge them by offering effective solutions at different cost levels.

## Author Note

The views expressed are those of the authors and do not reflect the official policy or position of the US Air Force, Department of Defense, or the US Government.

## References

- Ahuje, M. (2020) *How to Efficiently Detect Domain Generation Algorithms (DGA) in Kubernetes with Calico Enterprise | Tigera*. Available at: <https://www.tigera.io/blog/detecting-domain-generation-algorithms-dga-in-kubernetes/> (Accessed: 19 March 2021).
- Altigani, A., Hasan, S. and Barry, B. (2020) 'The Need for Polymorphic Encryption Algorithms: A Review Paper', *Journal of Theoretical and Applied Information Technology*, 15, p. 3. Available at: [www.ijatit.org](http://www.ijatit.org) (Accessed: 19 March 2021).
- Anthony, D. et al. (2012) 'A Behavior Based Covert Channel within Anti-Virus Updates', *Presentations and other scholarship*. Available at: <https://scholarworks.rit.edu/other/755> (Accessed: 7 May 2021).
- Bai, L. Y. et al. (2008) 'Covert channels based on jitter field of the RTCP header', in *Proceedings - 2008 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2008*, pp. 1388–1391. doi: 10.1109/IIH-MSP.2008.169.
- Bhaji, Y. (2007) *Understanding, Preventing, and Defending Against Layer 2 Attacks*. Available at: <https://www.menog.org/presentations/menog-6-7-8-9/Understanding, Preventing, and Defending Against Layer 2 Attacks.pdf> (Accessed: 5 May 2021).
- Cofer, R. (2019) *Sensitive Compartmented Information Facility (SCIF) and Special Access Program Facility (SAPF) Criteria*. Available at: [https://www.wbdg.org/FFC/NAVFAC/ATESS/navfac\\_far\\_east\\_scif\\_sapf\\_sept\\_2019\\_a.pdf](https://www.wbdg.org/FFC/NAVFAC/ATESS/navfac_far_east_scif_sapf_sept_2019_a.pdf) (Accessed: 5 May 2021).
- ComputerNetworkingNotes (2019) *Segmentation Explained with TCP and UDP Header*. Available at: <https://www.computernetworkingnotes.com/ccna-study-guide/segmentation-explained-with-tcp-and-udp-header.html> (Accessed: 19 March 2021).
- Convery, S. (2002) *Hacking Layer 2: Fun with Ethernet Switches*. Available at: <https://www.blackhat.com/presentations/bh-usa-02/bh-us-02-convery-switches.pdf> (Accessed: 5 May 2021).
- Coşar, M. and Karasartova, S. (2017) 'A firewall application on SOHO networks with Raspberry Pi and snort', in *2nd International Conference on Computer Science and Engineering, UBMK 2017*. Institute of Electrical and Electronics Engineers Inc., pp.
- Dakhane, D. M. and Tayde, J. H. (2018) 'Covert Channel and Countermeasures in the OSI Network Model', *International Journal of Advance Engineering and Research Development*, 5(03).
- Defense, D. of (1985) *Trusted Computer System Evaluation Criteria ["Orange Book"]*.
- Forcepoint LLC (2018) *How IPS engines and Layer 2 Firewalls inspect traffic*. Available at: <https://help.stonesoft.com/onlinehelp/StoneGate/SMC/6.3.0/GUID-AE26274B-EB56-4228-8B41-9EE90546A9A3.html> (Accessed: 7 May 2021).
- Gonçalves, R. A. S. (2011) *A MAC layer covert channel in 802.11 networks*. Naval Postgraduate School. Available at: <http://hdl.handle.net/10945/48138> (Accessed: 19 March 2021).
- Goodman, C. (2021) *SANS Institute Information Security Reading Room An Introduction to TEMPEST*.
- Imperva (no date) *OSI Model*. Available at: <https://www.imperva.com/learn/application-security/osi-model/> (Accessed: 19 March 2021).
- Li, Y. et al. (2011) 'Research based on OSI model', in *2011 IEEE 3rd International Conference on Communication Software and Networks, ICCSN 2011*, pp. 554–557. doi: 10.1109/ICCSN.2011.6014631.
- Liu, Z. et al. (2020) 'Covert Wireless Communication in IoT Network: From AWGN Channel to THz Band', *IEEE Internet of Things Journal*, 7(4), pp. 3378–3388. doi: 10.1109/JIOT.2020.2968153.
- Majidha Fathima, K. M. and Santhiyakumari, N. (2021) 'A Survey On Network Packet Inspection And ARP Poisoning Using Wireshark And Ettercap', in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. IEEE, pp. 1136–1141.
- MathWorks (2020) *802.11 MAC Frame Generation*. Available at: <https://se.mathworks.com/help/wlan/ug/802-11-mac-frame-generation.html;jsessionid=690f306477e61247887814369c1d> (Accessed: 19 March 2021).
- O'Neill, M. et al. (2017) 'TLS Inspection: How Often and Who Cares?', *IEEE Internet Computing*, 21(3), pp. 22–29. doi: 10.1109/MIC.2017.58.
- Park, W. and Ahn, S. (2017) 'Performance Comparison and Detection Analysis in Snort and Suricata Environment', *Wireless Personal Communications*, 94(2), pp. 241–252. doi: 10.1007/s11277-016-3209-9.
- Patsakis, C., Casino, F. and Katos, V. (2020) 'Encrypted and covert DNS queries for botnets: Challenges and countermeasures', *Computers and Security*, 88, p. 101614. doi: 10.1016/j.cose.2019.101614.
- Poornima, R. and Iswarya, R. J. (2013) 'An Overview of Digital Image Steganography', *International Journal of Computer Science & Engineering Survey*, 4(1), pp. 23–31. doi: 10.5121/ijcses.2013.4102.
- Prado, K. (no date) *GitHub - kelvins/steganography: Steganography: Hiding an image inside another*. Available at: <https://github.com/kelvins/steganography> (Accessed: 7 May 2021).

**Tristan Creek, Mark Reith and Barry Mullins**

- Saeli, S. *et al.* (2019) 'DNS Covert Channel Detection via Behavioral Analysis: a Machine Learning Approach', in *14th International Conference on Malicious and Unwanted Software*. Available at: <http://arxiv.org/abs/2010.01582> (Accessed: 5 May 2021).
- Salutari, F. (2018) *A closer look at IP headers*, *APNIC Blog*. Available at: <https://blog.apnic.net/2018/06/18/a-closer-look-at-ip-headers/> (Accessed: 19 March 2021).
- Sbrusch, R. (2006) *Network Covert Channels: Subversive Secrecy*. Available at: <https://www.sans.org/reading-room/whitepapers/covert/network-covert-channels-subversive-secrecy-1660> (Accessed: 19 March 2021).
- Schulzrinne, H. *et al.* (2003) *RTP: A Transport Protocol for Real-Time Applications*. Available at: <https://tools.ietf.org/html/rfc3550> (Accessed: 19 March 2021).
- Skybakmoen, T. (2018) *NEXT GENERATION FIREWALL COMPARATIVE REPORT Total Cost of Ownership (TCO)*. Available at: <https://www.fortinet.com/content/dam/fortinet/assets/analyst-reports/nss-labs-2018-ngfw-comparative-report-tco.pdf> (Accessed: 7 May 2021).
- Skybakmoen, T. and Dhanraj, M. (2017) *NEXT GENERATION FIREWALL COMPARATIVE REPORT Security Value Map™ (SVM)*. Available at: <https://www.esitechadvisors.com/wp-content/uploads/2018/06/Top-11-next-generation-firewalls-compared-June-6-2017>.
- Society, I. C. (2018) *802.3-2018 - IEEE Standard for Ethernet*, *IEEE*. Available at: <https://ieeexplore-ieee-org.afit.idm.oclc.org/document/8457469> (Accessed: 19 March 2021).
- Touch, J. (2013) *RFC 6864 - Updated Specification of the IPv4 ID Field*, *Internet Engineering Task Force*. Available at: <https://tools.ietf.org/html/rfc6864> (Accessed: 19 March 2021).
- Walsh, K. (2007) *Secure Facilities: Lessons from the SCIFs*, *CSO*. Available at: <https://www.csoonline.com/article/2121671/secure-facilities--lessons-from-the-scifs.html> (Accessed: 7 May 2021).
- Yadav, S., Jain, C. P. and Sharma, M. M. (2019) 'Smartphone Frequency Shielding with Penta-Bandstop FSS for Security and Electromagnetic Health Applications', *IEEE Transactions on Electromagnetic Compatibility*, 61(3), pp. 887–892. doi: 10.1109/TEM.2018.2839707.

# Deepfake Video Detection

Shankar Bhawani Dayal and Brett van Niekerk

University of KwaZulu-Natal, South Africa

[sbdayal15@gmail.com](mailto:sbdayal15@gmail.com)

[vanniekerkb@ukzn.ac.za](mailto:vanniekerkb@ukzn.ac.za)

DOI: 10.34190/EWS.21.110

**Abstract:** Deepfakes pose a threat to many aspects of society, such as election manipulation, involuntary pornography and fraud by means of identity theft. This paper aims to determine if deepfake models which are pre-trained on older datasets are still able to accurately detect whether a video is real or a deepfake from a newer dataset. From a comprehensive literature review, two papers were selected to be tested. The first model tested was from Afchar et al. (2018), was unable to run due to an error involving Keras, to no fault of the code. The model that was successfully tested on a sub-dataset was the XceptionNet model from the FaceForensics++ paper by Rössler et al. (2019). It was shown that the XceptionNet model was not able to effectively detect deepfake videos, having a 51.31% classification accuracy on the subdataset, further analysis of the results showed that it only had a 13.16% accuracy when detecting deepfake videos and it had 89.47% accuracy when detecting real videos. As the methods which are used to create deepfake material improve, the previous work which has been done will need to be tested on the material created by the newer methods to determine if they are still effective at detecting deepfakes.

**Keywords:** deepfake, FaceForensics++, generative adversarial networks, Xceptionnet, deepfake detection challenge dataset

---

## 1. Introduction

Deepfake media, such as videos and images, are the greatest form of crime created from artificial intelligence (Smith, 2020). The idea of deepfakes was first proposed on Reddit by a user called “deepfakes” (Fink and Diamond, 2020; Sample, 2020), who uploaded videos of his work which happened to be deepfake pornographic material of famous Hollywood actresses. This eventually resulted in Reddit banning the subreddit “r/deepfakes” due to a policy change concerning involuntary pornography (u/landoflobsters, 2018).

Effective deepfake detection methods are required to combat the threats that deepfakes pose, and more importantly deepfake detection methods that can remain effective against newer deepfake creation methods and thus staying at least one-step ahead. Deepfakes are mostly created using Generative Adversarial Networks (GANs) (GoodFellow et al., 2020). GANs utilize two neural networks working against each other, where the “generator” network creates material trying to fool the second network, and the second network which is the “discriminator” attempts to spot the real or fake instances. This is the adversarial component of the network. The discriminator is essentially a classifier (Google Developers, 2019). The generator and discriminator learn from each other, since the generator's output is linked directly to the discriminator's input, and when backpropagation is occurring, the discriminator's prediction/output is used as a signal for the generator to update its weights (Google Developers, 2019).

### 1.1 Threats of deepfakes

Deepfakes pose a threat to many fields, such as politics/election manipulation, pornography, and financial fraud by means of identity theft. The use of deepfakes in election manipulation has already been seen in India in February 2020, where a deepfake video shows a politician criticizing the Delhi government at the time (Christopher, 2020; Jee, 2020). This deepfake video was distributed on WhatsApp and allegedly reached 15 million people (Christopher, 2020). Deepfake pornographic material consists of 96% of all online deepfake material, 99% of which consists of women who work in the entertainment industry (Paul, 2019). Deepfake pornographic material can be used for defamation and/or blackmail, and can be damaging to the victim. The creation and distribution of deepfake pornographic is already a crime in California, Virginia and Texas (Paul, 2019; Ruiz, 2020) and on 20th December 2019, President Trump signed the USA's first federal law in relation to deepfakes (Fink and Diamond, 2020). Deepfakes have already been used for financial fraud, which occurred in 2019 where a UK energy firm was fooled into sending \$240,000 to another company, who they believed was a legitimate Hungarian supplier (Damiani, 2019; Stupp, 2019).

## **1.2 Aims and objectives**

The aim of this project is to validate the claimed accuracy of Rössler et al. (2019) and Afchar et al. (2018) by testing their pretrained models on a subdataset from the DFDC dataset (Dolhansky et al., 2020) available on Kaggle (2019). The belief is that the quality of the deepfakes in their datasets are not of high quality and therefore their results are high since the deepfakes are easy to detect. There is no doubt whether the models achieved the accuracy that is claimed (Afchar et al., 2018; Rössler et al., 2019); however, will these models achieve similar accuracies when tested on another deepfake dataset? Since these papers propose models that are able to detect deepfakes with a high accuracy, they should work on all types of deepfakes, which is what this project aims to validate.

The chosen evaluation method is classification accuracy, as in a commercial environment, the performance of an implemented prediction system will be measured by its classification accuracy. Prediction systems will do a tally of how many frames in the video were classified as real, and if that tally divided by the total number of frames is over a certain percentage, then the prediction will be real. This project aims to investigate how effective the deepfake models by Rössler et al. (2019) and Afchar et al. (2018) are when using different datasets. Rössler et al. (2019) and Afchar et al. (2018) used public datasets and are cited by many other papers (293 and 255 citations on Google Scholar, respectively); hence, this is a major factor in choosing to validate their results.

## **1.3 Paper structure**

The paper is structured as follows; a literature review is presented, which gives a detailed explanation of existing deepfake models as well as existing datasets. The methodology section outlines the work done by this project. This is followed by a presentation of the results and discussion, and lastly conclusion and future work.

## **2. Literature review**

There exists a variety of Deepfake Detection methods such as Faceforensics++ (Rössler et al., 2019) and MesoNet (Afchar et al., 2018). The summarised method of Faceforensics++ and MesoNet is using a face detection method to capture the faces from a picture or each frame of a video and then using a pre-trained Convolutional Neural Network (CNN) model to predict if that picture/frame contains a deepfake face/altered face. There are new novel detection methods such as eye blinking patterns (Jung, Kim and Kim, 2020).

Rössler et al. (2019) created a dataset called Faceforensics++ as well as another dataset in partnership with Google and Jigsaw (<https://jigsaw.google.com/>), which is a unit of Google that works on solutions to tackle threats to society. The FaceForensics++ dataset was created with four types of methods, namely Face2Face (Thies et al., 2016), FaceSwap (MarekKowalski, 2018), Deepfakes (Deepfakes, 2020) and Neural Textures (Thies et al., 2019). Face2Face is a system that transfers the expressions of source video (source face) to a target video (target face) while maintaining the target's face. Essentially Face2Face puts the expressions from one person onto another person's face. FaceSwap is a "graphic-based approach" which transfers the face region from a source video to a target video (MarekKowalski, 2018). The deepfake method used in Rössler et al. (2019) is a method available on GitHub (Deepfakes, 2020). Thies et al. (2019) propose a method utilising neural textures, which are learned feature maps that are trained during the scene capturing process. They performed testing on a variety of existing models (Afchar et al., 2018; Bayar and Stamm, 2016; Chollet, 2017; Deepfakes, 2020; Rahmouni et al., 2017). The best performing model is XceptionNet (Chollet, 2017), they fine-tuned the model and trained and tested it on their dataset and achieved an accuracy of 99.26% of raw video footage cropped on the face, and 82.01% accuracy when tested on the raw full video footage (not cropped on the face).

In Afchar et al. (2018), two CNN models to detect Deepfakes were created. The first network, Meso4, has four layers, and the second network is MesoInception-4. MesoInception-4 differs from Meso4 by replacing the first two layers with a variant of the Inception model (Szegedy et al., 2017). In Afchar et al. (2018), they created their own dataset and they used an existing dataset. The deepfake dataset they created and the existing dataset is Face2Face (Thies et al., 2016). They tested both models on the Face2Face dataset at different compression levels. The first network, Meso4, achieved 89.1% accuracy on the Deepfake dataset and 94.6% accuracy for the Face2Face classification score at 0 Compression level, 92.4% accuracy for the Face2Face classification score at 20 Compression level and 83.2% accuracy for the Face2Face classification score at 40 Compression level. The second network MesoInception-4 achieved 91.7% accuracy on the Deepfake dataset and 96.8% accuracy for the Face2Face classification score at 0 Compression level, 93.4% accuracy for the Face2Face classification score at

20 Compression level and 81.3% accuracy for the Face2Face classification score at 40 Compression level (Afchar et al., 2018).

In Jung, Kim and Kim (2020), they propose a method, called DeepVision, to detect if a video is a deepfake or not by analyzing the eye blinking patterns, while taking into account other variables such as age, gender, time of day, type of activity (static or dynamic). Currently they do not have an automated system to gather these variables, they have to manually gather these variables by watching the video. In Jung, Kim and Kim (2020), they utilize two algorithms to capture the eye blinking, namely Fast-HyperFace (Ranjan, Patel, and Chellappa, 2019) algorithm and EAR (Eye-Aspect-Ratio) algorithm (Soukupova and Cech, 2016). The Fast-HyperFace algorithm is good at detecting faces but not adequate at detecting eye blinking, and the EAR algorithm is good at detecting eye blinking but not good at detecting faces. In Jung, Kim and Kim (2020), they tested DeepVision on eight videos, and it achieved an accuracy of 87.5%, seven out of eight videos were identified correctly.

Güera and Delp (2018) created a recurrent neural network to detect deepfake videos. The model uses a CNN for frame feature extraction and the convolutional long short-term memory (LSTM) network to perform sequence analysis on the extracted features. They assembled their dataset, consisting of 600 videos, where 300 real/pristine videos came from HOHA dataset (Hollywood Human Actions) (Laptev, 2020) and the 300 deepfake videos came from undisclosed sources/websites. Güera and Delp (2018) tested their network against 20, 40 and 80 frames from each video. They state that they have 96.7%, 97.1% and 97.1% accuracy respectively when they tested it.

**Table 1:** Results of previous deepfake detection models

Paper	Model	Dataset	Accuracy/PerformanceMeasure
Rössler, et al (2019)	XceptionNet	FaceForensics++	82.01% accuracy on rawuncropped video footage
Afchar, et al (2018)	Meso4	Deepfakes	89.1% accuracy at 0compression,
		Face2Face (Thies et al., 2016)	94.6% accuracy at 0 compression level, 92.4% accuracy at 20 compressionlevel and 83.2% accuracy at40 compression level.
	Mesoinception-4	Deepfakes	91.7% accuracy at 0compression,
		Face2Face (Thies et al., 2016)	96.8% accuracy at 0 compression level, 93.4% accuracy at 20 compressionlevel and 81.3% accuracy at40 compression level.
Jung, S. Kim and K. Kim(2020)	DeepVision	DeepVision Dataset (8 videos)	87.5% accuracy
Güera and Delp (2018)	RNN	HOHA (Laptev, 2020) + undisclosedsources	96.7% accuracy with 20 frames per video, 97.1% accuracy with 40 frames pervideo and 97.1% accuracy with 80 frames per video

A recent challenge, Deepfake Detection Challenge (DFDC), on Kaggle (2019) finished recently which was hosted by AWS, Facebook, Microsoft, the Partnership on AI’s Media Integrity Steering Committee (Lyons, 2019), the winner receives \$500,000 and the four runner-ups receive a total of \$500,000. The winner of the competition achieved a log loss score of 0.42798 or rather he achieved an accuracy of 65.18% (Ferrer et al., 2020; Seferbekov, 2020). The discrepancy between the accuracy of published models (Afchar et al., 2018; Güera and Delp, 2018; Jung, Kim and Kim, 2020; Rössler et al., 2019) and a winner of a worldwide competition with 2265 teams, begs the question of how accurate these systems truly are. The caveat is that these models did not have that many datasets available to them at the time considering that the concepts of deepfakes are still fairly new in the academia world. As a result, they had to create their own datasets, as seen in Rössler et al. (2019) and Afchar et al. (2018), and the ways that deepfakes are being created are only growing in performance and quality.

The DFDC dataset (Dolhansky et al., 2020) uses five methods to create the dataset, namely DFAE (Deepfake Autoencoder), MM/NN (Morphable Mask/Neural Network) face swap, Neural Talking Heads (NTH), Face Swap GAN (FSGAN) and StyleGAN. DFAE is how most of the deepfakes seen on the internet are made, they are seen

in off-the-shelf products/software such as the DeepFaceLab (iperov, 2020). MM/NN face swap performs face swaps with a custom frame-based morphable-mask model. MM/NN face swap works by computing the facial landmarks in the source image and the target image and the pixels from the source face/image are morphed to match the landmarks in the target image, the method was adapted from Huang and De La Torre (2012). This technique works best when both faces (target and source) have similar expressions, otherwise the resultant video will have obvious discontinuities in the face, seeing as how most deepfake detection methods detect on a frame-by-frame basis, this would not affect the prediction, but this would be evident to a human observer. To overcome the discontinuities that occur they used a nearest-neighbours approach on the frame landmarks in order to find the best source/target face pair, this means that not every frame is created with the same two faces. Zakharov et al. (2019) create deepfakes using a GAN architecture, it utilizes two training stages, the first training stage is a meta-learning stage and the second training stage is a fine-tuning stage. FSGAN is another model that uses a GAN architecture to create deepfakes for face swapping and re-enactment from a source video/face to a target video/face, while accounting for facial expressions and the pose of the individuals (Nirkin, Keller, and Hassner, 2019). They modified the StyleGAN model (Karras, T., Laine, S. and Aila, 2019) to fit their purpose to produce face swaps between a given fixed identity descriptor onto a video by projecting this descriptor on the latent face space, they did this for every frame of the video. Dolhansky et al. (2020) use the most methods to create the dataset, compared to the four methods used by Rössler et al. (2019) and the two methods by Afchar et al. (2018).

### **3. Methodology**

The computer used for this project has:

- Operating System: Windows 10
- Processor: Intel(R) Core (TM) i5-7300HQ CPU @ 2.5GHz
- Graphics Card: Nvidia GeForce GTX 1050
- Installed RAM: 8GB
- System Type: 64bit Operating System, x64-based processor

#### **3.1 Creating the dataset**

The first step is creating the sub-dataset from the DFDC dataset. Since the DFDC is fairly large in size (471.84 GB zipped), the dataset is also segmented into smaller files ( $\approx$  10GB zipped) to allow for smaller downloads. The first four segmented datasets were chosen, "00.zip", "01.zip", "02.zip" and "03.zip". Each zipped folder contains a set of videos, real and fake, as well as a metadata.json file which lists the name of the video, the label of the video ("REAL" or "FAKE"), and if the video is fake, the source/original video. The "03.zip" folder got corrupted and due to data issues, the decision was made not to redownload the folder.

Since each video in the file is not labelled real or fake in the title, the next step is to separate them into two folders, a "real" folder and a "fake" folder. Two folders are created for each segmented dataset, i.e there are six total folders. A python program is created to open the metadata.json file that is in each segmented folder, create a list of the fake videos present in the folder by selecting each video where label equals "FAKE" in the metadata.json file. This list of fake videos now has the name of each video present in the segmented folder, in a for-loop for each video in the list:

- Add "\\" and the name of the video to the end of the file path of the segmented folder, this gives you the file path of the video (source path)
- Create a new destination path with the the file path of the output folder + "\\"+video name (destination path)
- Check if the source path exists, and if the source path exists, move the file from the source path to the destination path.

This is done for real videos in the same manner, and we repeat this for each segmented folder. A new folder was created called "Test" folder, which contains 150 videos total, 50 videos from each segmented folder, where 25 videos are real and 25 videos are fake. They are stored in either a "real" folder or "fake" folder. The reason for only testing 150 videos is due to computation time. Each frame in a video is checked for a face, and if a face is

found, the frame is then sent to the model where a prediction is returned, this is the reason for why the computation time is high.

The testing for Rössler et al. (2019) and Afchar et al. (2018) was implemented in Spyder (<https://www.spyderide.org/>) on Anaconda (<https://www.anaconda.com>).

### **3.2 Setting up the environments and adapting the code to the dataset**

To run Rössler et al. (2019) and Afchar et al. (2018), a separate environment has to be created for each implementation. This is because they both have different requirements, such as Rössler et al. (2019) runs on Python 3.6 and Afchar et al. (2018) runs on Python 3.5.

The environment created for Afchar et al. (2018) is called “Meso”. The requirements are listed on Github (Afchar, 2018). Where possible, conda -install was used to install the requirements otherwise pip install was used. There are only 6 listed requirements for Afchar et al. (2018):

- Python 3.5 (<https://www.python.org/>)
- Numpy 1.14.2 (<https://pypi.org/project/numpy/1.14.2/>)
- Keras 2.1.5 (<https://faroit.com/keras-docs/2.1.5/>)
- Imageio (<https://pypi.org/project/imageio/>)
- FFMPEG (<https://www.ffmpeg.org/download.html>)
- face\_recognition (Geitgey, 2018)

The recommended installation method for the face\_recognition method is to use pip install (Geitgey, 2018). The installation was not successful due to certain packages not being compatible. The next step taken to try and install face\_recognition was to use conda install ([https://anaconda.org/conda-forge/face\\_recognition](https://anaconda.org/conda-forge/face_recognition)), the installation was successful with no errors. The package version of face\_recognition was not specified on the requirements list (Afchar, 2018).

The first attempt at launching Spyder in the Meso environment with the model and attempting to run the code resulted in a crash of Spyder. The only fix was to do a full reset of Spyder, thereafter the Meso environment was deleted, and a new environment was created, also called Meso, with the same list of requirements. Spyder successfully launched with no crashes or bug errors.

The first environment created for Rössler et al. (2019) was called “FaceForen”, and the requirements are quite extensive and can be found on Github along with the code (Rössler, 2020). The first attempt to install the requirements, involved using the pip command “pip install -r requirements.txt”, it stopped installing the requirements once it could not successfully install a package, thereafter manually installing the requirements occurred and there were issues with installing 3 packages:

- mkl-fft==1.0.10
- mkl-random==1.0.2
- torch==1.0.1.post2

The first two packages did not appear to exist. The error happened to be a naming error; it should have been:

- mkl\_fft==1.0.10
- mkl\_random==1.0.2

This is an error with the requirements list on Github (Rössler, 2020). The torch package version was not available on the official pytorch website (<https://pytorch.org/get-started/previous-versions/>), a pip installation of torch==1.1.0 was successful.

After all packages were installed, the next step was to launch Spyder to test the model, and subsequently there were errors with the installed packages. The decision was made to create a new environment called “ffv2” and to try to find the closest compatible versions for every listed requirement.



In the new environment, ffv2, each requirement was manually installed; the closest torch version found is pytorch==1.0.1. The difference in the package name is because “torch” is the package name when using pip install, and “pytorch” is the package name when using conda install. The package installation instructions for older versions of pytorch is from the official pytorch website (<https://pytorch.org/get-started/previous-versions/>). After installing each package on ffv2, a test launch of Spyder shows that the installation of each package is successful.

On testing the Afchar et al. (2018) model on a test video, an unsolvable error occurred, “ERROR (theano.gof.opt) : Optimization failure due to: local\_abstactconv\_check”, which is caused by the package Theano (<https://github.com/Theano/Theano>) and the fix for this problem is to update to a later version of keras (Lamblin, 2017). Theano is a required package that is installed when installing keras. Therefore, this project did not get Afchar et al. (2018) model to work. The testing for Rössler et al. (2019) went without any problems. The code needed slight adaptations; it had no evaluation metric for the output/prediction video.

When predicting videos from the “real” folder from the Test folder:

- A counter is kept for each frame predicted as REAL,  $r$
- A counter is kept for the number of frames  $n$
- *Classification Accuracy for video* =  $(r/n) \times 10$
- A counter is kept for every video that is predicted REAL,  $tR$ , i.e. where the *classification accuracy for video*  $\geq 50$
- *Classification Accuracy for real videos* =  $(tR/75) \times 100$  there are 75 real videos in the “real” folder.

When predicting videos from the “fake” folder from the Test folder:

- A counter is kept for each frame predicted as FAKE,  $f$
- A counter is kept for the number of frames  $n$
- *Classification Accuracy* =  $(f/n) \times 100$
- A counter is kept for every video that is predicted as FAKE,  $tF$ , ie where the *classification accuracy for video*  $\geq 50$
- *Classification Accuracy for fake videos* =  $(tF/75) \times 100$ , there are 75 fake videos in the “fake” folder.

The classification accuracy for each video and the subdatasets are written to a file. To test Rössler et al. (2019), their modified XceptionNet model is being used with the pretrained weights, trained on the FaceForensics++ dataset. The pretrained weights are available on the Github page (Rössler, 2020), the pretrained weights used is “full\_raw.p” for the XceptionNet model.

#### 4. Results and discussion

Afchar et al. (2018) was not able to be tested, to no fault of the code or the model, this is an error with Keras, as such, there is no way we can validate or verify the effectiveness of the model on a portion of the DFDC dataset.

Rössler et al. (2019) with pretrained weights achieved 13.1578947368% accuracy for detecting deep fakes and 89.4736842105% for detecting real videos, with an overall accuracy of 51.3157894736%.

**Table 2:** Table showing results achieved for Rössler et al. (2019)

	Pretrained XceptionNet tested on DFDC dataset (this papers results)	XceptionNet results from Rössler et al. (2019)
<b>Classification accuracy for deepfake videos:</b>	13.1578947368%	Unknown
<b>Classification accuracy for realvideos:</b>	89.4736842105%	Unknown
<b>Average classification accuracy:</b>	51.3157894736%	82.01% accuracy on raw uncropped video footage

**Table 3:** Table showing the worst 10 detections of deepfake videos by XceptionNet tested on the DFDC subdataset

Name of the video:	Classification accuracy:
aaknzywids.mp4	0.0% fake
aasmohwrt.mp4	0.0% fake
aayrffkzxn.mp4	0.0% fake
abxtdjyru.mp4	0.0% fake
afnxnrrqsj.mp4	0.0% fake
aheocfkxjx.mp4	0.0% fake
aijltdlrj.mp4	0.0% fake
akfjqoantp.mp4	0.0% fake
akpuczgfpk.mp4	0.0% fake
alqiqhnrza.mp4	0.0% fake

**Table 4:** Table showing the best 10 detections of deepfake videos by XceptionNet tested on DFDC subdataset

Name of the video:	Classification accuracy:
aejvkfbtxs.mp4	97.56944444444444% fake
aimkjacvip.mp4	94.66192170818505% fake
ablzpwqhcc.mp4	88.62068965517241% fake
ambabjrbt.mp4	87.58389261744966% fake
ajeegjyzk.mp4	82.0% fake
alxodlppci.mp4	78.76712328767124% fake
adckadazdl.mp4	67.90123456790124% fake
ahofrimoni.mp4	59.51417004048582% fake
acdckfsyev.mp4	57.89473684210527% fake
agdivudslh.mp4	54.66666666666664% fake

As we can see a portion of the results, based on only 150 videos, their modified XceptionNet model, was not able to detect the deepfake videos in the DFDC dataset. The believed reason for the classification accuracy for the deepfakes being so low is that the FaceForensics++ dataset, is not a true generalization of deepfakes (Dolhansky et al., 2020), the FaceForensics++ dataset contains only 1000 videos or about about half a million edited images (Rössler, 2020). In relation, the DFDC dataset has 128,154 videos. These results shows that models trained on the first generation of deepfakes, is not as accurate as they claim to be. The model architecture is not in question, but rather the quality of the dataset.

**Table 5:** Table showing worst 10 detections of real videos by XceptionNet tested on DFDC subdataset

Name of the video:	Classification accuracy:
chfkrpvgnz.mp4	19.014084507042252% real
chqqxfuuzi.mp4	27.66666666666668% real
apedduehoy.mp4	30.33333333333336% real
awkvatcshx.mp4	31.103678929765888% real
bwtyeopljx.mp4	31.1787072243346% real
eppyqpgewp.mp4	47.0% real
bvsnqubtjc.mp4	49.831649831649834% real
fopjiyiqd.mp4	50.0% real

Name of the video:	Classification accuracy:
fsaronfupy.mp4	51.162790697674424% real
eyguqfmgzh.mp4	54.333333333333336% real

**Table 6:** Table showing the best 10 detection of real videos by XceptionNet on DFDC subdataset

Name of the video:	Classification accuracy:
aayrffkzxn.mp4	100.0% real
almnlnfyu.mp4	100.0% real
bvpeerislp.mp4	100.0% real
cxsvvnxyz.mp4	100.0% real
cxwcpdspni.mp4	100.0% real
dvwpvqdflix.mp4	100.0% real
dzrrklwrgn.mp4	100.0% real
exseruhiuk.mp4	100.0% real
extbidoov.mp4	100.0% real
exxqlfnpbz.mp4	100.0% real

## 5. Conclusion

This project has successfully tested a popular and well respected paper by Rössler et al. (2019) and found the accuracy is potentially not as high as it is claimed to be, the believed reason for the accuracy being so low is that the deepfakes in the Faceforensics++ dataset are firstly, not big enough in terms of the amount of videos present in the dataset, secondly, does not accurately represent each scenario in which a deepfake video may occur. For example, the videos in the DFDC dataset had a variety of physical scenarios which involves the target person in the video (both real and deepfake videos), such as a person pacing back and forth in a room, a person sitting upright and looking straight at a camera and so on. The Faceforensics++ dataset chose videos where the targets face is front facing, thus it would be expected for any model trained on only front facing deepfakes to not be able to accurately or effectively detect deepfake videos when the target is not front facing. The Faceforensics++ dataset is considered as a first-generation dataset (Dolhansky et al., 2020). The methods used to create the deepfakes in the Faceforensics++ dataset are also seen in other papers such as in Afchar et al. (2018), the premise is that they will have deepfakes of the same quality which would not adequately train them to detect deepfakes that are created with methods seen today.

This project shows the evident need for constant testing of previous work to determine if and when they are no longer sufficiently adequate at detecting deepfake material, which indicates that the fight against deepfakes may be an unending task.

This project can be improved by testing more models, which are trained on datasets other than Faceforensics++, which this project did not do. This will illustrate if deepfake models which are trained on older datasets are in fact unable to reliably and effectively determine whether a video/picture is real or a deepfake. The Faceforensics++ dataset is approximately 3.5TB in size, and therefore it was not downloaded due to lack of storage space. This project can be trained on the DFDC dataset and tested on the Faceforensics++ dataset to determine if the accuracy achieved is greater than trained on the Faceforensics++ dataset and tested on the DFDC dataset, as this project accomplished. The effectiveness of a deepfake detection model seems to be highly dependent on the quality of the dataset, this might suggest that deepfake detection models may always be proved to be ineffective once a newer deepfake creation method is used, since methods which utilize the GAN architecture may utilize a deepfake detection model as the discriminator to create the new generation of deepfake material.

## References

- Afchar, D. (2018) MesoNet, Github, [online], <https://github.com/DariusAf/MesoNet>.  
 Afchar, D., Nozick, V., Yamagishi, J. and Echizen, I., (2018) "Mesonet: a compact facial video forgery detection network", *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, Hong Kong, China, 1-7.

- Bayar, B. and Stamm, M.C., (2016) "A deep learning approach to universal image manipulation detection using a new convolutional layer", *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*, 5-10.
- Chollet, F., (2017) "Xception: Deep Learning with Depthwise Separable Convolutions," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 1800-1807.
- Christopher, N., (2020) "We've Just Seen the First Use of Deepfakes in an Indian Election Campaign", *Vice*, 18 February, [online], accessed 4 November 2020, <https://www.vice.com/en/article/igedjb/the-first-use-of-deepfakes-in-indian-election-by-bjp>.
- Damiani, J., (2019) "A Voice Deepfake Was Used to Scam a CEO Out of \$243,000", *Forbes*, 3 September, [online], accessed 4 November 2020, <https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=74bb412d2241>.
- Deepfakes (2020) Faceswap, Github, [online], accessed 5 November 2020, <https://github.com/deepfakes/faceswap>.
- Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M. and Ferrer, C.C., (2020) "The deepfake detection challenge dataset", arXiv, [online], <https://arxiv.org/abs/2006.07397>.
- Ferrer, C., Dolhansky, B., Pflaum, B., Bitton, J., Pan, J. and Lu, J., (2020) "Deepfake Detection Challenge Results: An Open Initiative to Advance AI", Facebook AI, [online], <https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai/>.
- Fink, D. and Diamond, S., (2020) "Deepfakes: 2020 and Beyond", *The Recorder*, 3 September [online], accessed 4 November 2020, <https://www.law.com/therecorder/2020/09/03/deepfakes-2020-and-beyond/?slreturn=20201116162814>.
- Geitgey, A. (2018) face\_recognition, Github, [online], [https://github.com/ageitgey/face\\_recognition](https://github.com/ageitgey/face_recognition).
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., (2020) "Generative adversarial nets", *Communications of the ACM* 63(11), 139-144.
- Google Developers. (2019) "Overview of GAN Structure", *Generative Adversarial Networks*, [online], accessed 4 November 2020, [https://developers.google.com/machine-learning/gan/gan\\_structure](https://developers.google.com/machine-learning/gan/gan_structure).
- Güera, D. and Delp, E.J., (2018) "Deepfake Video Detection Using Recurrent Neural Networks," *15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Auckland, New Zealand, 1-6.
- Huang, D. and De La Torre, F., (2012) "Facial action transfer with personalized bilinear regression", *12th European Conference on Computer Vision*, Proceedings Part II Florence, Italy, 144-158.
- iperov (2020) Deepfacelab, Github, [online], <https://github.com/iperov/DeepFaceLab>.
- Jee, C., (2020) "An Indian Politician is Using Deepfake Technology to Win New Voters", *MIT Technology Review*, 19 February, [online], accessed 4 November 2020, <https://www.technologyreview.com/2020/02/19/868173/an-indian-politician-is-using-deepfakes-to-try-and-win-voters/>.
- Jung, T., Kim, S. and Kim, K., (2020) "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern", *IEEE Access* 8 83144-83154.
- Karras, T., Laine, S. and Aila, T., (2019). "A style-based generator architecture for generative adversarial networks", *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, 4396-4405.
- Kaggle. (2019) Deepfake Detection Challenge, [online], <https://www.kaggle.com/c/deepfake-detection-challenge>.
- Lamblin, P., (2017) "MILA and the future of Theano", theano-users Google Groups, 28 September, [online], accessed 5 November 2020, <https://groups.google.com/g/theano-users/c/7PqQ8BZutbY/m/rNClfvAEAwAJ>.
- Laptev, I., (2020), Learning Human Actions from Movies, Département d'Informatique, ENS, [online], <https://www.di.ens.fr/~laptev/actions/>.
- Lyons, T., (2019) "The Partnership on AI Steering Committee on AI and Media Integrity", The Partnership on AI, 5 September, [online], <https://www.partnershiponai.org/the-partnership-on-ai-steering-committee-on-ai-and-media-integrity/>.
- MarekKowalski (2018) Faceswap, GitHub, [online], accessed 5 November 2020, <https://github.com/MarekKowalski/FaceSwap/>.
- Nirkin, Y., Keller, Y. and Hassner, T., (2019) "FSGAN: Subject agnostic face swapping and re-enactment", *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, South Korea, 7183-7192.
- Paul, K., (2019) "California Makes 'Deepfake' Videos Illegal, But Law May be Hard to Enforce", *The Guardian*, 7 October, [online], accessed 4 November 2020, <https://www.theguardian.com/us-news/2019/oct/07/california-makes-deepfake-videos-illegal-but-law-may-be-hard-to-enforce>.
- Rahmouni, N., Nozick, V., Yamagishi, J. and Echizen, I., (2017) "Distinguishing computer graphics from natural images using convolution neural networks", *2017 IEEE Workshop on Information Forensics and Security*, Rennes, France, 1-6.
- Ranjan, R., Patel, V.M., and Chellappa, R., (2019) "HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41(1), 121-135.
- Rössler, A., (2020) FaceForensics++, Learning to Detect Manipulated Facial Images, Github, [online], <https://github.com/ondyari/FaceForensics>.
- Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J. and Niessner, M., (2019) "Faceforensics++: Learning to detect manipulated facial images", *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, South Korea, 1-11.
- Ruiz, D., (2020) "Deepfakes Laws And Proposals Flood US", Malwarebytes Labs, 23 January, [online], accessed 4 November 2020, <https://blog.malwarebytes.com/artificial-intelligence/2020/01/deepfakes-laws-and-proposals-flood-us/>.

**Shankar Bhawani Dayaland Brett van Niekerk**

- Sample, I., (2020) "What Are Deepfakes – And How Can You Spot Them?", *The Guardian*, 13 January, [online], accessed 2 November 2020, <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them>.
- Seferbekov, S., (2020) dfdc\_deepfake\_challenge, Github, [online], [https://github.com/selimsef/dfdc\\_deepfake\\_challenge](https://github.com/selimsef/dfdc_deepfake_challenge).
- Smith, A., (2020). "Deepfakes Are The Most Dangerous Crime of the Future, Researchers Say", *The Independent*, 5 August, [online], accessed 2 November 2020, <https://www.independent.co.uk/life-style/gadgets-and-tech/news/deepfakes-dangerous-crime-artificial-intelligence-a9655821.html>.
- Soukupova, T., and Cech, J. (2016) "Real-time eye blink detection using facial landmarks", *21st Computer Vision Winter Workshop*, Rimske Toplice, Slovenia, 42-50.
- Stupp, C., (2019) "Fraudsters Used AI to Mimic CEO'S Voice in Unusual Cybercrime Case", *The Wall Street Journal*, 30 August, [online], accessed 4 November 2020, <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>.
- Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A., (2017) "Inception-v4, inception-resnet and the impact of residual connections on learning", *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17)*, 4278–4284.
- Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C. and Niessner, M., (2016) "Face2face: Real-time Face Capture and Reenactment of RGB videos", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2387-2395.
- Thies, J., Zollhöfer, M. and Niessner, M., (2019) "Deferred neural rendering: Image synthesis using neural textures", *ACM Transactions on Graphics* 38(4), 1-12.
- u/landoflobsters, (2018) Update On Site-Wide Rules Regarding Involuntary Pornography And The Sexualization Of Minors, [online], accessed 2 November 2020, [https://www.reddit.com/r/announcements/comments/7vxzrb/update\\_on\\_sitewide\\_rules\\_regarding\\_involuntary/](https://www.reddit.com/r/announcements/comments/7vxzrb/update_on_sitewide_rules_regarding_involuntary/).
- Zakharov, E., Shysheya, A., Burkov, E. and Lempitsky, V., (2019) "Few-shot adversarial learning of realistic neural talking head models", *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, South Korea, 9458-9467.

# A Shoestring Digital Forensic Cyber Range for a Developing Country

Jaco du Toit and Sebastian von Solms  
University of Johannesburg, South Africa

[jacodt@uj.ac.za](mailto:jacodt@uj.ac.za)

[basievs@uj.ac.za](mailto:basievs@uj.ac.za)

DOI: 10.34190/EWS.21.008

**Abstract:** The 2020 Covid-19 lockdown forced many universities and other training institutions to rethink how training occurs. Training typically consists of theoretical aspects, but they may also contain practical elements. Practical computer training typically requires computer hardware and software of a certain standard, configuration, and setup. Such practicals then make use of dedicated lab equipment and/or virtual environments. The non-contact requirement in the Covid-19 lockdown shifted the focus towards totally online virtual environments. In some settings, such virtual environments are also known as cyber ranges. Many organisations provide access to pre-built cyber ranges, and individuals can quickly join a cyber range. Existing cyber ranges offer individuals access through free or pay accounts. Organisations can contract cyber range service providers to use their platforms for teaching and learning. The National Institute for Standards and Technology (NIST) also has a draft cyber range guide that organisations can use to plan and set up their cyber range. Setting up a cyber range for an organisation's purposes requires significant investment and skills. The research described in this paper flows from the problems during the Covid-19 pandemic to offer specific courses which need particular lab environments. Students had to have access to specialised software programs and large data sets to do their practicals. Lecturers spent a large amount of time helping students, having unreliable Internet connections and low-level hardware devices, to do the required practicals. The problem resulted in a decision to plan, implement, and maintain such a cyber range for a typical South African context. The resultant cyber range solution had to allow students with the limitations mentioned above to perform their practical work properly. The main limitation was a shoestring budget. The shoestring cyber range is also compared against more feature-rich environments to highlight the features that are not included in such a shoestring cyber range. The research concludes that providing basic cyber range functionality may be sufficient in specific circumstances.

**Keywords:** cyber security, cyber range, digital forensics, virtual

---

## 1. Introduction

Tertiary educational institutions had to adapt their regular contact lectures during the 2020 Covid-19 pandemic (Sari and Nayir, 2020, Lassoued, Alhendawi & Bashithalshaaer, 2020). Some courses required both theoretical and practical teaching aspects. Many different mechanisms were implemented to address the theoretical part of teaching during which non-contact instruction had to continue. The theoretical education shifted towards online learning (Mishra, Gupta & Shree, 2020).

Practical experience, required by many courses, was a challenge for tertiary educational institutions. Institutions used videos that demonstrated practical skills, but the general feeling was that there is no substitution for the real thing (Azlan, et al., 2020). Teaching that involved the practical use and application of computer programs had lower entrance barriers for online education. Still, in a developing country such as South Africa, even this was not as simple. There are considerable inequalities in South Africa, with many students not having Internet access or computing devices or even a stable electricity supply (Jandrić, et al., 2020, Motala and Menon, 2020).

Having the computing devices with Internet access also did not always prove to be without problems. Not all computing devices have the correct requirements for practical exercises. The software required by students may require special licenses, which are not available to all students. Specific configuration and data files may also require students to utilise the software required by the practical exercises fully. In a normal situation, the department would have solved these challenges by giving students access to lab computers that already have the correct software installed, configured, and licensed.

One solution that addresses the software, configuration and licensing aspect is using a virtual computing environment. The virtual environment is preconfigured with the software and data. Students are granted access to the virtual environment through an Internet connection. In cyber security, the virtual computing environments are also known as test beds or cyber ranges. The term 'cyber ranges' is used throughout this article to refer to both.

The challenge with software ranges is that you either need a significant amount of in-house infrastructure to host it yourself or have enough funds for a hosted solution. The research presented in this paper describes the architectural components and implementation of a cyber range developed using minimal funds and resources.

The rest of the paper is organised according to the research methodology followed. Section 2 gives the context in which the cyber range was developed, and it highlights the challenges in providing practical training in a locked-down environment. Section 3 introduces some of the essential aspects of cyber ranges, some of the use cases it is used for, how somebody may implement it, and the general components in these cyber ranges. Section 4 describes how the solution to the problem described in section 2 was designed and implemented, using minimal resources. Section 5 discusses the success and constraints of the shoestring cyber range. Section 6 provides some general comments on the research conducted and some of the areas for future research.

## **2. Background and problem**

Traditional practical training for a digital forensics course requires students to complete several exercises in a controlled lab environment. The digital forensics course teaches students the process and methods used to find forensically sound evidence on computers in cyber crime or cyber security breaches. The digital forensics course allows students to find hypothetical evidence inside pre-captured computer images. Students have a well-established theoretical and practical knowledge in conducting a digital forensic investigation at the end of the digital forensics course. The knowledge and skills acquired in the course help them be more productive as part of the digital forensic investigation team.

The lab environment consists of several physical desktop computers making use of specialised software. The exercises conducted by students consist of them having to find specific data items in a simulated scenario. The data items are hidden on the actual physical computers. Still, in most cases, the students need to interrogate a data file, a forensically captured image of a target computer.

During the lockdown, the first problem was that students did not have the necessary software available to start with the exercises. The courses' lecturers had to change their approach and provide access to either evaluation versions of the software or free, open-source versions of the software.

As soon as the software was made available, the second challenge was that not all students could easily install the software on typical operating systems available on student computers. Some software requires unique configuration and support to make them functional.

The third problem experienced was that data files could be, on average, about 1GB in size. The size is a significant amount of data for students to download over limited or costly Internet connectivity.

One solution that addresses the three problems is to create a virtualised lab environment with the necessary software already installed and the data files pre-loaded. However, several challenges need to be addressed before such a virtual environment can be made available to students. Some of the challenges are:

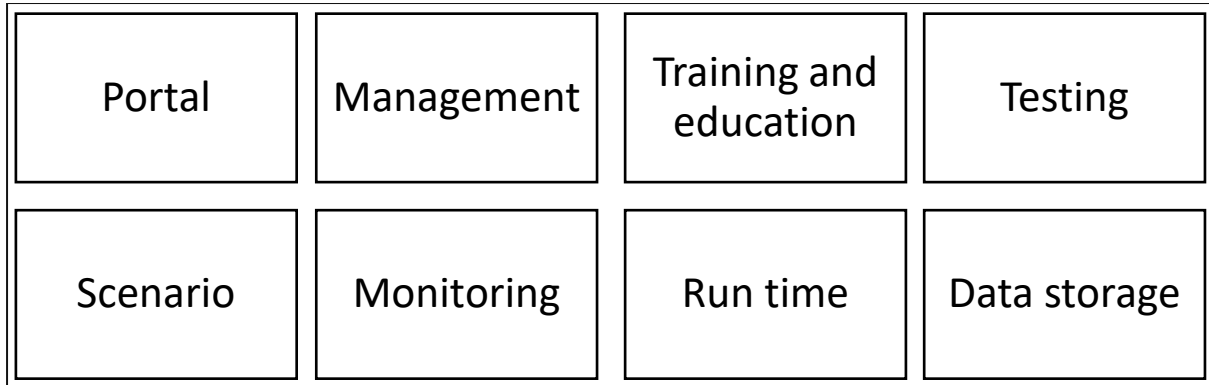
- 1. The environment must allow only one student access to a computing environment.
- 2. The computing environment limits the time available for a student to work in the computing environment.
- 3. The computing environment needs to be reset after each student completed their work.
- 4. The system must allow lecturers to update and modify the computing environments with minimum effort.
- 5. The system must allow students to connect to the computing environment without the need for specialised software.

The list above lists required features for the shoestring cyber range described in Section 4 of the document. In-contact practical exercises traditionally require physical lab computers. In an environment where teaching had to move towards online learning, using a virtual lab environment allows students to complete their practical exercises. The virtual lab environment is also known as a test bed or cyber range. A basic overview of typical cyber ranges is discussed in the next section.

### 3. A brief overview of cyber ranges

Cyber ranges are used for teaching, but it also plays a role in experimentation and research. Several authors claim that the first cyber range made available for public use was the National Cyber Range developed by the Defense Advanced Research Projects Agency (DARPA) (Ferguson, Tall & Olsen, 2014, Ranka, 2011, Ukwandu et al., 2020).

Yamin, Katt, and Gkioulos (2020) describes eight (8) high-level functional architectural components associated with most cyber ranges. The eight parts are listed in **Figure 1**.



**Figure 1:** Functional architecture of a cyber range (Yamin, Katt & Gkioulos, 2020)

A brief overview of the eight functions are (Yamin, Katt & Gkioulos, 2020):

- Portal. The portal acts as the interface between the cyber range and the various users. The users include students, admins, teachers, and many more.
- Management. Management describes the management of resources but also the administration of user roles.
- Training and education. The training and education function provides a tutoring system, keeps track of the students' scores during scenarios, and analyses the student's work after an exercise.
- Testing. The testing module takes a test case and transforms it into a scenario. After scenario activities, the results from the test are also evaluated in this functional area.
- Scenario. The scenario module allows various scenarios to be created, edited, and deployed.
- Monitoring. The execution of a scenario may be monitored and analysed by different users of the system.
- Run time. The cyber range computing environment runs on physical, virtual, or hybrid environments. The run time describes the various physical and virtual components and how they interact with each other.
- Data storage. The data storage acts as a repository for all tests, logs, virtual machine images, software, and anything else that is either deployed to or extracted from the run time environment.

A full spectrum literature study is too vast to include in this research. However, given the literature cited in this section, the functional architecture described by Yamin, Katt & Gkioulos (2020) provided the researchers a well-established basis to start the design for the intended shoestring cyber range. The next section describes a shoestring cyber range architecture and compares it with the list of functional components described in **Figure 1**.

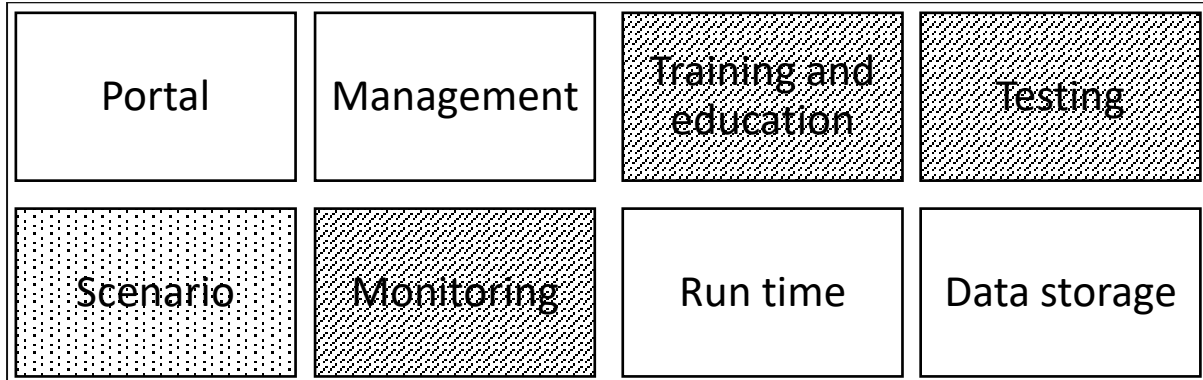
### 4. A shoestring cyber range architecture

Designing a cyber range to be useful for a developing country, like South Africa, with a minimal budget, requires that some of the functional aspects of a typical cyber range, defined by Yamin, Katt & Gkioulos (2020), either be left out, scaled-down or be done manually. This section describes the functional architecture as it relates to the functional architecture discussed in section 3. This section also describes how the functional architecture is implemented.

The researchers decided that several functional architectural components were either left out or minimised according to the original definitions to reduce the time it took to design and implement the shoestring cyber



range. **Figure 2** describes three levels of implementation, explicitly highlighting well-executed (clear blocks), basic (lightly contrasted blocks), and minimal (dark contrasted blocks) implementation levels. The clear blocks describe areas that the researchers implemented reasonably wholly. These areas include the Portal, Management, Run time, and Data storage. The Scenario function, filled with dots, giving it a light contrast, describes fundamental functionality with a lot of manual intervention by the lecturer and administrator. The blocks filled with diagonal black lines, giving it a dark contrast, indicate no implementation. These include Training and education, Testing, and Monitoring.



**Figure 2:** Functional architecture of shoestring cyber range

The general application of the shoestring cyber range allows students to connect to a specific virtual machine remotely. The cyber range server controls the virtual machine. The cyber range server ensures specific scenarios are made available to the virtual machine used by a student. Each of the implemented functions is described in more detail in the next sub-sections.

#### 4.1 Portal

The portal functionality provides access to the lecturer, student, and administrator. The portal is accessible through a web interface, with some of the admin functionality available through command line terminal functionality.

The portal is implemented as a basic PHP application that changes the configuration datastore depending on whether the user chooses to start, stop, or connect to a virtual machine. The command-line interface is necessary for the administrator to update the raw operating system images, override specific config settings when essential, and modify scenarios associated with the different machines.

#### 4.2 Management

The shoestring cyber range's management functionality is responsible for ensuring that virtual machines start, stop, and reset to their initial states when required. It manages each student's time allocation to ensure a student can only use a computing environment for the allocated time frame. It also addresses each computing environment's access control and provides only students with the correct password can access the computing environment.

#### 4.3 Scenario

The scenario component's functionality allows the lecturer to control what software is made available to each computing environment. It ensures that software is installed, users are created, and general operating system configuration is applied to the computing environment. It also makes specific data files and software available to the computing environment.

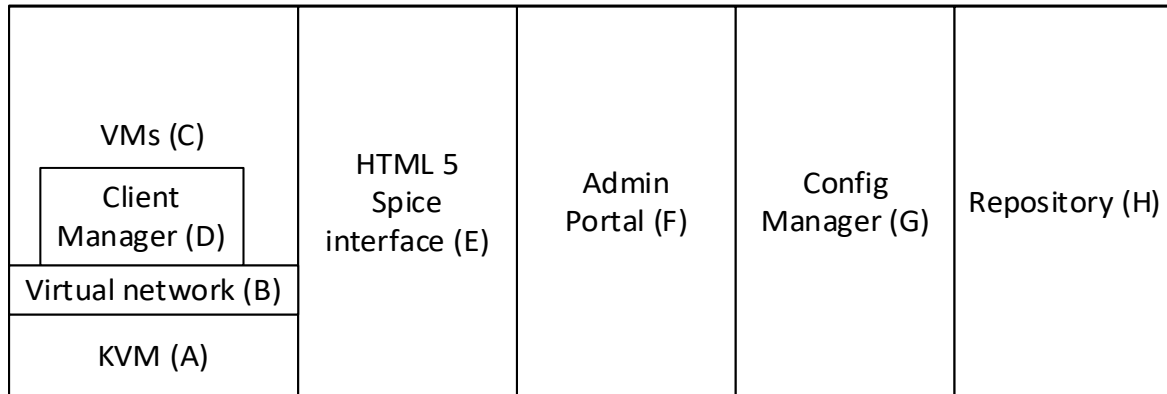
Changing the scenarios, such as creating new forensic data files, happens outside the cyber range's current implementation. The capturing of forensic images needs to occur outside the implemented cyber range design. Future versions of the shoestring cyber range may include this functionality, but this is one of the areas removed to minimise the resource requirements for the cyber range, thereby minimising the costs of the implementation.

#### 4.4 Data storage

The data storage functionality ensures that virtual machine images can be stored on the computer. The data storage functionality provides that software and forensic data files are made available to virtual machines for scenario exercises.

#### 4.5 Run time

The run time functionality provides the infrastructure required by the cyber range. This section describes how the cyber range is implemented using the components depicted in **Figure 3**.



**Figure 3:** Implementation architecture

The whole system is implemented on one server. The server has 30GB of RAM, with 2TB of storage space implemented on 10 000 RPM SCSI Disks, implemented as RAID 1 and RAID 5 volumes. **Figure 3** provides a basic overview of the implementation architecture of the shoestring cyber range. Each of the most relevant components in **Figure 3** is described below:

- The student-accessible virtual machines run on a Kernel-based Virtual Machine hypervisor (KVM). The primary server services (Components E – H) run as separate components on the host server, but the student-allocated virtual machines (C) are implemented on the KVM hypervisor.
- The virtual machines are allocated a virtual network on the host. The network is a private connection that uses a Network Address Translation (NAT) bridge to access the Internet. Students do not access the virtual machines directly through the network but instead, use the HTML5 Spice interface (E). The virtual network's role is to ensure the virtual machine can communicate with the server and have access to the Internet.
- The implemented virtual machines are currently running Windows 10 operating systems since they need to use the required operating system for the specific courses. It is also possible to implement different guest operating systems, depending on the requirements of the lecturers.
- On each virtual machine, a particular Client Manager (D) component communicates with the Config Manager (G) on the server. The Client Manager is responsible for installing any assigned software, creating users, copying files, and updating the configuration of the Virtual Machine for the student according to settings defined by the lecturer. The Client Manager runs during the virtual machine startup and connects to the server every five minutes to see any new tasks assigned to it. The Client Manager is written as a Python script to ensure cross-operating system functionality.
- Students connect to the virtual machine using an HTML 5 Spice (Red Hat Inc., NA) connection. The Spice protocol allows the virtual machine screen to be streamed to the student and allows the student keyboard and mouse remote control capabilities in their assigned virtual machine. The student does not need to install any special software on their computers and only requires an Internet connection while doing their assigned exercises.
- The admin portal allows the administrator to make some fundamental changes to the cyber range. Some functionality is available to the admin through a web interface, while another more specialised functionality requires the admin to connect using a command-line interface.
- The Config Manager consists of two services. The first service manages the running of the virtual machines. Depending on the configuration, it ensures that virtual machines start when they are scheduled to run and

ensure that virtual machines are reset when the time allocated to a student runs out. The second service manages connections from the Client Manager (D) and sends instructions that must be executed on the virtual machines.

- The repository contains all the virtual machine images, the software required for the designed scenarios, the forensic images, and the Config Manager's configuration files. The virtual machine images are stored as copy-on-write (COW) virtual machines-based files that allow functionality such as snapshots to be created for the virtual machines. The configuration files required by the Config Manager are implemented as JavaScript Object Notation (JSON) files.

The shoestring cyber range can be described according to five of the nine functional cyber architecture components defined by Yamin, Katt & Gkioulos (2020). The run time functional feature that describes how the cyber range is implemented consists of eight different elements. The run time components describe how the shoestring cyber range is implemented using virtual machines, specialised streaming protocols, server-side and client-side parts, and data and configuration repositories.

An implementation goal tried to minimise the level of custom development necessary in the shoestring cyber range. Many of the components described in this section used built-in operating system functionality or open-source features. The parts that required some custom coding were in the Admin Web Portal (F), the Client Manager (D), and the Config Manager (G). The implementation of the shoestring cyber range occurred over about 40 days. Throughout the design, development, and implementation period, the implementation team acquired many new skills, notably in PHP, Python, and KVM virtual machines' configuration. The next section describes the testing on the Shoestring cyber range and the results of the tests produced.

## 5. Testing and results

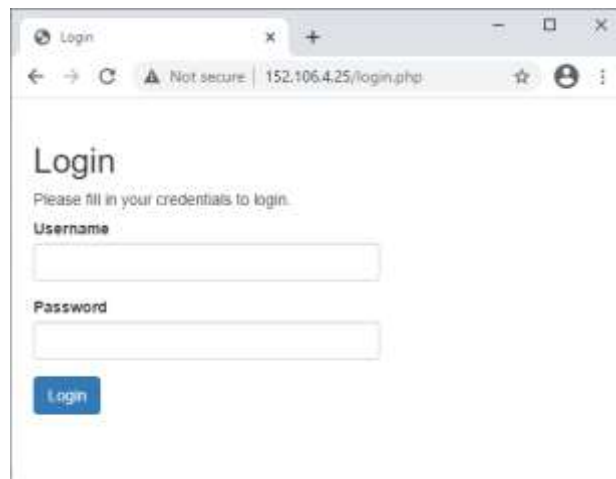
The shoestring cyber range went through an operational testing phase. The tests conducted were based on the challenges identified in section 2. The tests are summarised in **Table 1**.

**Table 1:** Functional tests

Test Number	Test Description.
1	Student access control to a virtual machine.
2	Student time limit implementation.
3	Resetting each virtual machine before a new student can use it.
4	Modifying the virtual machine configuration with minimal effort.
5	Remote connection to computing environment without specialised software.

### 5.1 Test 1: Student access control

The implementation team developed an elementary access control and login system for the web console. The login system ensures that only registered account holders can access the end-user functionality of the system. **Figure 4** displays the basic login page to ensure students have a valid username and password before gaining access to the system.



**Figure 4:** Login page for basic access control

### 5.2 Test 2: Student time limit

Each virtual machine created on the server has a specific configuration entry stored in JavaScript Object Notation (JSON). When a virtual machine starts, the Config Manager (G in Figure 3), records the starting time in the 'starttime' configuration entry. It also records a 'deadline' value. The deadline value is the number of minutes since the virtual machine's start that the virtual machine can run. The Config Manager uses the two entries to determine when the Config Manager should destroy the virtual machine.

Figure 5 shows all the configuration items for each virtual machine with some values. Figure 5 shows some example values for 'starttime' and 'deadline'. The value of 'starttime' in Figure 5 is '2020-12-07T12:37:37.608121', and the 'deadline' value is '60'.

```

{
  "lock" : "false",
  "assigned": "true",
  "client": "2",
  "port" : "5960",
  "deadline": "60",
  "reboot": "true",
  "runningstate": "off",
  "vmpass": "2",
  "software": [
    {
      "command": "C:\\SOFTWARE\\osf\\osf.exe /SILENT /LOADINF=C:\\SOFTWARE\\osf\\OSFSetup.inf",
      "testFile": "c:\\program files\\osforensics\\osforensics.exe",
      "testSize": 471328
    }
  ],
  "starttime": "2020-12-07T12:37:37.608121"
}

```

Figure 5: Virtual machine configuration settings

The test team started a virtual machine for the test. Figure 6 shows that virtual machine number 1 is running and started at 10:10 on 12 January 2020. It further indicates a 60-minute run duration limit. The test team left the machine running for just over 60 minutes, and the test team confirmed that the virtual machine was not running anymore, and the values displayed on the admin page, was reset.

Number	Status	Start time	Run duration	Actions
1	running	2021-01-12T10:10:20.282033	60	Stop Connect
2	off	0	0	Start Connect
3	off	0	0	Start Connect

Figure 6: Admin page after the test team started the virtual machine.

### 5.3 Test 3: Reset each virtual machine before use

The virtual machine received a snapshot after creating each virtual machine. The snapshot ensures that the system can revert changes made to the virtual machine. The snapshot allows students to make changes to running images without affecting any other student that might use the same virtual machine at a different time slot.

The Config Manager (G in **Figure 3**) ensures that whenever a virtual machine stops, that there are several steps performed. The first step is to revert the virtual machine image to the original snapshot. It then stops the virtual machine and does a few other actions, such as resetting the start time and deadline configuration items (described in section 5.2).

The test was done by connecting to a virtual machine and then creating a file on the desktop. The virtual machine was then stopped and started again. After the test team restarted the virtual machine, the file created was no longer visible on the virtual machine. The Config Manager log (**Figure 7**) also confirms that the snapshot is reverted when the virtual machine is destroyed. The record shows that at 11:10:29, the Config Manager reverted the snapshot image. At 11:10:50, the Config Manager destroyed the running virtual machine. The rest of the entries show that the 'starttime' and 'deadline' config entries were reset.

```
2021-01-12 11:10:29,468 virsh snapshot-revert --domain win10-1 --snapshotname win10-1_snapshot1 --running
2021-01-12 11:10:50,693 virsh snapshot-revert --domain win10-1 --snapshotname win10-1_snapshot1 --running:0
2021-01-12 11:10:50,693 virsh destroy win10-1 Domain win10-1 destroyed
2021-01-12 11:10:51,012 virsh destroy win10-1:0
2021-01-12 11:10:51,013 Client starttime:0
2021-01-12 11:10:51,013 Client deadline:0
2021-01-12 11:10:51,013 Client runningstate:off
```

Figure 7: Extract of Config Manager log

#### 5.4 Test 4: Modifying the virtual machine configuration with minimal effort.

The administrator can modify the configuration of a virtual machine in several ways. One way requires the administrator to remove a virtual machine's snapshot, changing some configuration on the virtual machine and then retaking the snapshot. The method gives the administrator full control over what needs to happen on the virtual machine. Still, it is very cumbersome to implement on several virtual machines since the changes must occur on each image.

Another option available to the administrator is to execute instructions on the virtual machines by modifying the Config Manager's configuration settings (G in **Figure 3**). The software configuration items describe various programs or scripts that the Client Manager executes on the virtual machine client (software in **Figure 5**). The Client Manager (D in **Figure 3**) receives a list of software entries from the Config Manager. Each entry consists of the necessary command to install the software and information to test if the Client Manager installed the software successfully. The software configuration item is an array of configuration settings that consists of three parts:

- **command.** The command entry describes the script, application, or any command that needs to be executed on the virtual machine. Example: The command entry in **Figure 8** is a Windows command that resets the Engineer user account on the virtual machine to '123'. It further creates a file called UserStat.chk on the virtual machine. The UserStat.chk file is only created if the Client Manager successfully changed the password.
- **testFile.** The Client Manager uses the testFile to determine if it should execute the command. The example displayed in **Figure 8** looks for the existence of UserStat.chk. If the file exists, it further evaluates the testSize.
- **testSize.** The testSize indicates the size of the testFile in bytes on the file system. If the testFile entry's size is not correct, then the Client Manager assumes that the Client Manager should execute the command again.

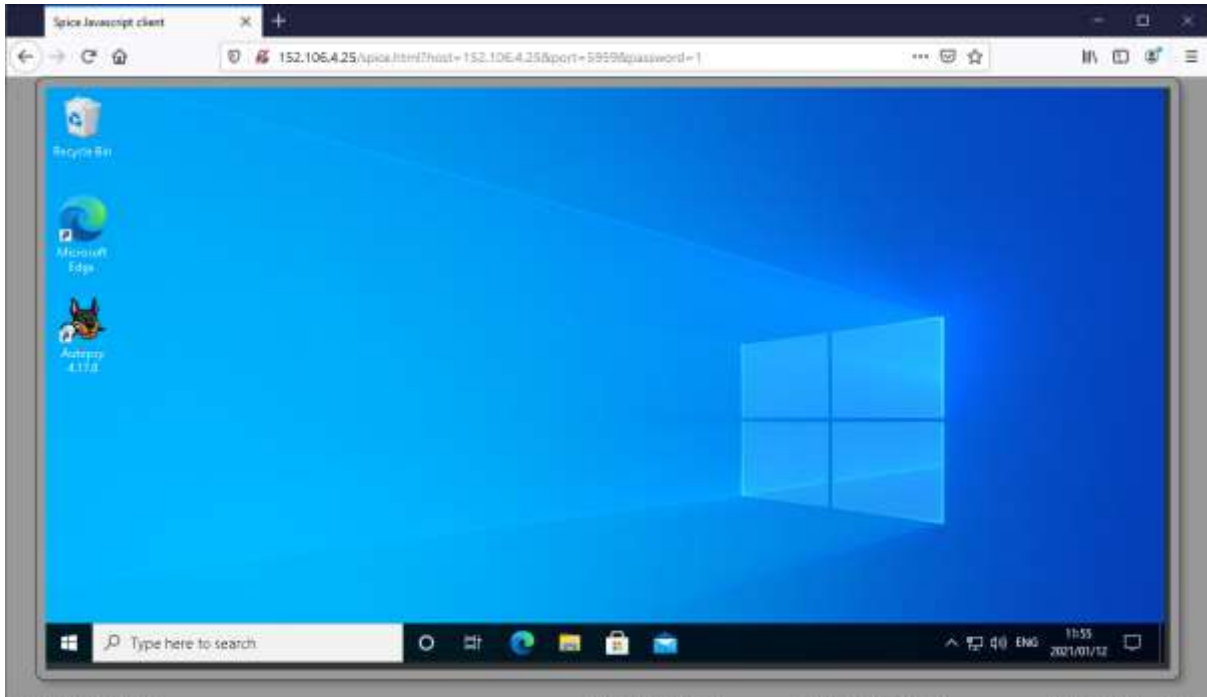
```
{
  "command": "cmd /c net user Engineer 123 & echo User Password Reset > c:\\windows\\temp\\UserStat.chk",
  "testFile": "c:\\windows\\temp\\UserStat.chk",
  "testSize": 22
},
```

Figure 8: Configuration entry for resetting the engineer user account's password

The test was conducted using the same setting described in **Figure 8** but to reset the Engineer user's password to '1234'. The test team left the virtual machine for five minutes to ensure that the Client Manager had an opportunity to execute any new configuration instructions. The Client Manager successfully changed the Engineer's password to '1234'.

### 5.5 Test 5: Remote connection to computing environment without specialised software.

The remote connection test ensures that students can connect to assigned virtual machines with minimal software deployed to their computers. Connection to the web interface was tested using Google Chrome, Firefox, and Microsoft Edge browsers. Each of the Internet browsers successfully connected to the virtual machine. **Figure 9** is a screenshot of the Firefox browser connected to the virtual machine which runs Windows 10.



**Figure 9:** Firefox internet browser connecting to a virtual machine

## 6. Conclusion and future research

The shoestring cyber range is a proof-of-concept cyber range used by students. The research team will add features to the cyber range to address more functionality by the lecturers, admins, and students over time. One of the immediate features that require attention is students' ability to start their virtual machines if they accidentally initiated a shutdown.

The research team will also extend the admin user interface to provide better scheduling capabilities and allow lecturers to publish instructions for practical exercises on the existing web interface without reverting to secondary sources for instructions.

The advantage of the shoestring cyber range is that the admins and lecturers have full control of the environment and provides a basis to extend functionality as they see fit. One of the most significant drawbacks to the shoestring cyber range is that it requires consistent administration and support. The in-house administration and support may incur extra costs in the long run to the organisation.

Future research and tests for the shoestring cyber range will highlight the costing aspects, specifically regarding pure cloud vs. on-premises deployment and admin costs.

The Shoestring cyber range demonstrated that a cyber range could be implemented using a limited budget and limited time for a typical South African-based university.

## References

- Azlan, C.A., Wong, J.H.D., Tan, L.K., A.d. Huri, M.S.N., Ung, N.M., Pallath, V., Tan, C.P.L., Yeong, C.H. & Ng, K.H. (2020) "Teaching and learning of postgraduate medical physics using internet-based e-learning during the COVID-19 pandemic – A case study from Malaysia", *Physica medica*, Vol 80, pp 10-16.

### **Jaco du Toit and Sebastian von Solms**

- Ferguson, B., Tall, A. & Olsen, D. (2014) *National Cyber Range Overview, 2014 IEEE Military Communications Conference*, IEEE.
- Jandrić, P., Hayes, D., Truelove, I., Levinson, P., Mayo, P., Ryberg, T., Monzó, L.D., Allen, Q., Stewart, P.A., Carr, P.R., Jackson, L., Bridges, S., Escaño, C., Grauslund, D., Mañero, J., Lukoko, H.O., Bryant, P., Fuentes-Martinez, A., Gibbons, A., Sturm, S., Rose, J., Chuma, M.M., Biličić, E., Pfohl, S., Gustafsson, U., Arantes, J.A., Ford, D.R., Kihwele, J.E., Mozelius, P., Suoranta, J., Jurjević, L., Jurčević, M., Stekete, A., Irwin, J., White, E.J., Davidsen, J., Jaldemark, J., Abegglen, S., Burns, T., Sinfield, S., Kirylo, J.D., Kokić, I.B., Stewart, G.T., Rikowski, G., Christensen, L.L., Arndt, S., Pyyhtinen, O., Reitz, C., Lodahl, M., Humble, N., Buchanan, R., Forster, D.J., Kishore, P., Ozoliņš, J.J., Sharma, N., Urvashi, S., Nejad, H.G., Hood, N., Tesar, M., Wang, Y., Wright, J., Brown, J.B., Prinsloo, P., Kaur, K., Mukherjee, M., Novak, R., Shukla, R., Hollings, S., Konnerup, U., Mallya, M., Olorundare, A., Achieng-Evensen, C., Philip, A.P., Hazzan, M.K., Stockbridge, K., Komolafe, B.F., Bolanle, O.F., Hogan, M., Redder, B., Sattarzadeh, S.D., Jopling, M., SooHoo, S., Devine, N. & Hayes, S. (2020) "Teaching in the age of covid-19", *Postdigital science and education*, Vol 2, (3):pp 1069-1230.
- Lassoued, Z., Alhendawi, M. & Bashitialshaer, R. (2020) "An exploratory study of the obstacles for achieving quality in distance learning during the COVID-19 pandemic", *Education sciences*, Vol 10, (9):
- Mishra, L., Gupta, T. & Shree, A. (2020) "Online teaching-learning in higher education during lockdown period of COVID-19 pandemic", *International Journal of educational research open*, pp 100012.
- Motala, S. and Menon, K. (2020) "In search of the 'new normal': Reflections on teaching and learning during covid-19 in a South African university", *Southern African review of education*, Vol 26, (1): pp 80-99.
- Ranka, J. (2011). *National cyber range*. DEFENSE ADVANCED RESEARCH PROJECTS AGENCY ARLINGTON VA STRATEGIC TECHNOLOGY.
- Red Hat Inc. (NA). "SPICE", [online], <https://www.spice-space.org/index.html>.
- Sari, T. and Nayır, F. (2020) "Challenges in distance education during the (covid- 19) pandemic period.", *Qualitative research in education (2014-6418)*, Vol 9, (3): pp 328-360.
- Ukwandu, E., Farah, M.A.B., Hindy, H., Brosset, D., Kavallieros, D., Atkinson, R., Tachtatzis, C., Bures, M., Andonovic, I. & Bellekens, X. (2020) "A review of cyber-ranges and test-beds: Current and future trends", *arXiv preprint arXiv:2010.06850*.
- Yamin, M.M., Katt, B. & Gkioulos, V. (2020) "Cyber ranges and security testbeds: Scenarios, functions, tools and architecture", *Computers & Security*, Vol 88

# A Strategy for Implementing an Incident Response Plan

Alexandre Fernandes<sup>1</sup>, Adail Oliveira<sup>1,2</sup>, Leonel Santos<sup>1,2</sup> and Carlos Rabadão<sup>1,2</sup>

<sup>1</sup>School of Technology and Management, Polytechnic of Leiria, Portugal

<sup>2</sup>Computer Science and Communication Research Centre, Polytechnic of Leiria, Portugal

[alexandre.fernandes@ipleiria.pt](mailto:alexandre.fernandes@ipleiria.pt)

[adail.oliveira@ipleiria.pt](mailto:adail.oliveira@ipleiria.pt)

[leonel.santos@ipleiria.pt](mailto:leonel.santos@ipleiria.pt)

[carlos.rabadao@ipleiria.pt](mailto:carlos.rabadao@ipleiria.pt)

DOI: 10.34190/EWS.21.080

**Abstract:** With the exponential growth of the Internet, several challenges and security threats arise. Those threats are due to the lack of adequate security mechanisms, security policy flaws, increasing usage of mobile devices, mobility, and user's naivety. Although organisations try their best to deploy effective security solutions and practices, there will always be security incidents. Therefore, they must place detection methods to identify those threats and vulnerabilities. On the other hand, response activities must be established to deal with and respond to the detected incidents. An Incident Response Plan (IRP) aims to provide an organisation with an easy-to-follow guide that leads to a quick and effective incident response. The implementation of such a plan is not an easy task. To implement an IRP requires an organisation a lot of research and analysis of the existing frameworks and examples. Most frameworks explain how to set up a Computer Security Incident Response Team and how they should handle incidents, but only a few instruct how to implement a plan. The proposal of this paper is to present a practical strategy on how to implement an IRP, complementing the existing incident response frameworks, thus reducing the difficulty of creating an effective and useful plan. The study and proposal of this topic come from the research and experience developed during the implementation of an academic Security Operation Centre. The paper starts by presenting the most relevant incident response frameworks and related work. It then proposes a flexible strategy for creating an IRP that can be adjusted to any organisation's scope and objectives. During the strategy presentation, the various domains of incident response are presented. Finally, strategies for its implementation will be introduced. As the main contribution of this work, the reader will be able to understand the common structure and content of an IRP and to create their own plan.

**Keywords:** incident response, CSIRT, framework, cybersecurity, security operations centre

---

## 1. Introduction

As technology grows and evolves, so does cybercrime. Attackers have increasingly sophisticated tools and strategies, and most organisations simply cannot prevent or block them. With the increase of computer attacks, companies are beginning to recognise the need and criticality of protecting their organisations, information and businesses from security threats or incidents.

After the occurrence of an incident, it is the security incident response team (CSIRT) that is responsible for handling the situation. However, for the team to respond adequately to a security incident, it is necessary to have some previously defined procedures. One of the most important procedures is the IRP, which is the focus of this work. The IRP is a procedural piece that describes the actions that an organisation must take to deal with security incidents and, consequently, minimise their impact, whether from cyber-attacks, data breaches, policy violations or other security incidents. As a reference document, its content must be practical and easy to understand. However, due to the diversity of needs and organisations' limitations, it is impossible to have a single plan that suits all organisations. The creation of an IRP is a complex task, and despite the diversity of existing standards and frameworks to guide this task's execution, organisations still do not know how to structure or implement theirs.

Some of the difficulties experienced by organisations are the lack of sufficient knowledge and the diversity of standards and frameworks, which is the cause of the delay in implementing a good IRP.

As the main contribution of this work, the reader will be able to understand the typical structure and content of an IRP and to create their own plan. Past this section, this paper is organised as follows: Section 2 offers the related works, including the analysis of standards and frameworks, examples of IRP and other scientific articles. Section 3 provides and discusses the proposed strategy and its model, presenting the various model's domains. Section 4 offers paths to implementation of the discussed strategy and model. Section 5 concludes this paper and mentions future work. The study and proposal of this work come from the research and experience developed during the implementation of an academic Security Operation Centre.



## **2. Related work**

Initially, in this section, several reference standards and frameworks related to the implementation of an incident response program were collected and analysed to understand the recommendations proposed by the main reference entities and assess whether they are easily applicable or simple to follow. Subsequently, several examples of incident response plans, proposed by entities from the most diverse activity sectors, were studied, crossing their content with the previously analysed standards and frameworks. A search was also carried out in the leading databases of scientific works.

### **2.1 Standards and frameworks**

We sought to understand the main objectives and what type of information they provide by taking the most popular incident management frameworks and standards. Some of the analysed documents are: (i) SANS: Creating and Managing an Incident Response Team (Proffitt, Timothy; SANS, 2007); (ii) RFC 2350: Expectations for Computer Security Incident Response (Brownlee & Guttman, 1998); (iii) CERT: Handbook for Computer Security Incident Response Teams (CSIRTs) (West-brown, Stikvoort, & Kossakowski, 1999); (iv) NIST 800-61: Computer Security Incident Handling Guide (Scarfone, Grance, & Masone, 2012); (v) ENISA: Good Practice Guide for Incident Management (ENISA, 2010); (vi) ISACA: Incident Management and Response (ISACA, 2012); and (vii) ISO/IEC 27035-2:2016: Information Security Incident Management (International Organization for Standardization, 2016). The SANS document (i) includes the definition of the different CSIRT services, policies, and standards, identifies a sequence of phases regarding incident response and ends with some important details about CSIRT members. Regarding RFC 2350 (ii), this RFC provides CSIRT teams with a way to publicly publicise their services and scope in a comprehensive, simple, and common structure between different entities, thus, allowing CSIRT constituents to know its CSIRT's policies and procedures. Created by CERT, the Handbook for CSIRTs (iii) aims to provide materials and suggestions for the creation and operation of a CSIRT and an incident handling service. The document defines a framework that includes the definition of the mission, organisation, services, CSIRT policies, among others. In addition, it also addresses several details regarding functions and interactions during the handling of incidents. NIST 800-61 (iv) has a structure similar to the previous document. This document begins by defining how to organise the security incident response through the creation of policies, plans and procedures, as well as the structure of a CSIRT. Subsequently, it presents a framework of several phases of an incident, including a checklist and recommendations. Finally, it includes some topics on coordination and information sharing. The ENISA guide (v) mainly includes practical tips for application in the incident handling process, presenting a workflow and a life cycle for incident response. Like the previous ones, this document also addresses and offers suggestions for the definition of mission, organisation, and responsibilities in this matter. Finally, the topics of communication and cooperation between other CSIRTs or authorities are addressed, together with the issues of outsourcing and presentation to management. The ISACA document (vi) also defines an incident life cycle's phases, the associated security strategies, and other governance activities. This document recommends the use of the COBIT framework (IT Governance Institute, 2007) to structure incident response processes. Finally, ISO/IEC 27035-2 (vii) presents guidelines for planning and preparing for incident response. It mainly focuses on defining and updating security policies and a security incident management plan. It also has several topics about creating a CSIRT and its relationship with other organisations or entities. It ends with recommendations on the topics of awareness, training, IRP testing and on ways to evaluate and improve the different processes.

Almost all the presented frameworks include the definition of a workflow or a life cycle of a security incident with several recommendations on how to act during any of its phases. These documents present suggestions on implementing or designing a CSIRT and what to include in an IRP. However, there are no recommendations on how it should be structured, nor what path to follow when implementing it.

### **2.2 Examples of published Incident Response Plans**

Following the previous documents evaluation, several incident response plans from different activity sectors were identified and collected. To evaluate each plan's content, the authors considered the frameworks and standards to determine the most important topics to include in a plan. The identified topics are: Purpose and Scope; Roles and Responsibilities; IR Life Cycle; Communication Plan; Metrics; SLAs; Plan Test and Review; IR Training; Taxonomy; Incident Classification; and Playbooks. The performed analysis is systematised in Table 1.

**Table 1:** Comparison of several IRPs

Incident Response Plan	Purpose and Scope	Roles and Responsibilities	IR Life Cycle	Communication Plan	Metrics	SLAs	Plan Test and Review	IR Training	Taxonomy	Incident Classification	Playbooks
University Carnegie Mellon	✓	✓	✓	±	±	✗	±	±	✗	✗	✗
Tulane University	✓	✓	✓	✓	✗	✗	✗	✗	✗	✓	✓
University of Adelaide (Data breach)	✓	✓	✓	✓	✗	✗	✗	±	✗	✗	✓
University of Alabama	✓	✗	✗	✓	±	✗	✗	✗	✗	✗	✓
LGPD - PROCEMPA	✗	±	✓	✓	✗	✗	✗	✗	✗	✗	✗
Stinson Leonard Street	✓	✓	✓	✓	✗	✓	±	✗	✓	✓	✗
Criminal Justice Information Center	✓	✗	✓	±	✗	✗	±	✓	✗	✓	✓
OSCIO EIS (Sample)	✗	✓	✓	✓	✗	✗	±	±	✗	✗	✗
Universidade Federal do Rio de Janeiro	✓	✗	±	✗	✗	±	±	✗	✗	✗	✗
HU-UFJF (EBSERH)	±	✓	✓	✗	✓	✗	✗	✗	✓	✓	✗
EGPr-TIC	✓	✓	✓	✗	✓	✗	✗	✗	✗	✓	✗
Texas Southmost College	✓	✓	✓	✗	✗	✓	✗	±	✗	✓	✗
Western Oregon University (Data Breach IRP)	✗	✓	✗	✓	✗	✗	±	±	✗	✗	✓
JSM Marketing Services	✓	✓	✓	✓	✗	✗	✓	✗	✗	✗	✓
Total (✓/±/✗)	10/1/3	10/1/3	11/1/2	8/2/4	2/2/10	2/1/11	1/6/7	1/5/8	2/0/12	6/0/8	6/0/8

± - Reference to the topic or information too limited; ✓ - Information present; ✗ - Information not present

According to the table above, the most used topics are "Purpose and Scope", "Roles and Responsibilities", "IR Life Cycle", and "Communication Plan". There is some divergence in the parameters in the different plans analysed, which is probably due to: (i) the complexity and diversity of standards and frameworks; (ii) the specificities of each organisation; and (iii) the absence of systematisation of what an IRP should be. Therefore, it is understood that there is a need to design a practical IRP implementation strategy capable of contributing to this problem's resolution.

### 2.3 Scientific contributions

In addition to studying existing standards and frameworks and analysing existing IRPs, research for scientific papers related to this topic was also performed to find work that might assist in the practical implementation of an IRP. For this objective, a search was carried out in the leading databases for scientific works (IEEE, ACM,

others) produced in the last five years, using the terms considered most relevant, systematised in the following search key: "security AND incident AND response AND (framework OR plan OR methodology OR playbook)". As a result of this research, about eight dozen works were identified, namely in international conferences, journals, and books.

The most relevant works identified are: (i) SOTER: A Playbook for Cybersecurity Incident Management (Onwubiko & Ouazzane, 2020), that in addition to analysing the different terminologies used in the management of security incidents and providing a mapping of equivalences, proposes and discusses an incident classification and prioritisation scheme, to finally present a framework for incident response procedures according to its context, classification and time of occurrence; (ii) Towards the Development of an Integrated Incident Response Model for Database Forensic Investigation Field (Al-Dhaqm, Razak, Siddique, Ikuesan, & Kebande, 2020), that proposes suitable steps for constructing and integrating the Incident Response Model that can be relied upon the database forensic investigation field; and (iii) Context for the SA NREN Computer Security Incident Response Team (Mooi & Botha, 2016), that defines the business requirements for establishing a CSIRT, within the context of SA NREN CSIRT, resulting in "strategic" framework that sets a background for the establishment processes. The remaining articles focus on other subjects not relevant to this work.

Although the scientific papers presented similar address subjects, none of the reviewed documents provides a strategy or practical model for implementing an IRP, making this article an original contribution.

### 3. Strategy for designing an IRP

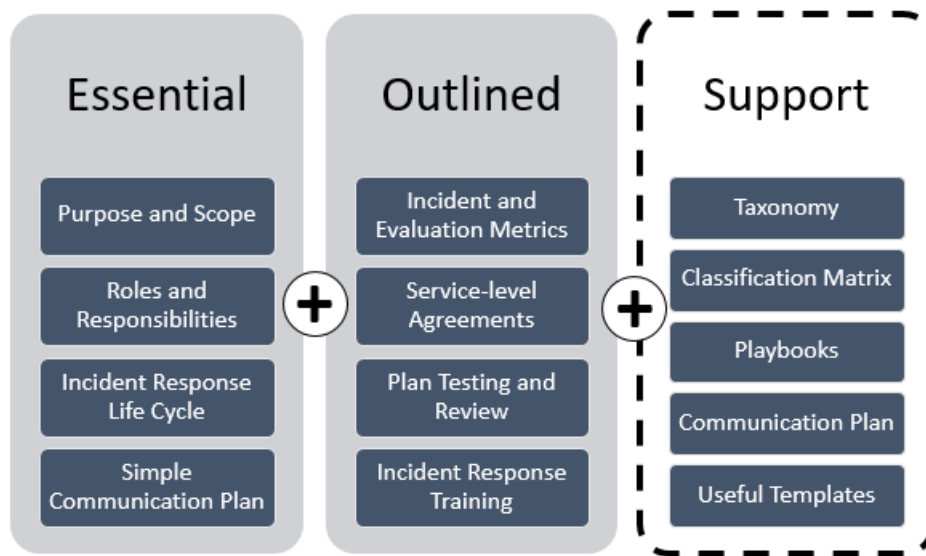
Following the lack of a practical IRP implementation strategy, the authors understood the need to contribute to the security incident response process's systematisation, having outlined an IRP implementation strategy for this purpose. At first, the potential difficulties and concerns in implementing an IRP were analysed to identify the strategy's objectives. Not having enough knowledge to implement an IRP, the existence of many standards and frameworks, and the delay in implementing a good plan are some of the difficulties experienced by organisations. Thus, considering these difficulties, the defined strategy should be phased and progressive, adjustable to different realities, and both practical and quick to implement. These requirements led us to the definition of a practical model for an IRP design.

The related work analysis and the experience acquired during the activities in an Information Security Office identified some relevant topics to consider in an IRP. Table 2 presents the relation between the specified subjects and the standards/frameworks that contemplate them. The ticked boxes mean that the respective document provides material or recommendations on implementing the subject.

**Table 2:** Topics coverage on frameworks

Framework/ Standard	Purpose and Scope	Roles and Responsibilities	IR Life Cycle	Communication Plan	Metrics	SLAs	Plan Testing and Review	IR Training	Taxonomy	Incident Classification	Playbooks	Templates
SANS: CM IRT	-	X	X	-	-	-	-	X	X	X	-	-
IETF: RFC 2350	X	-	-	-	-	-	-	-	X	X	-	-
CERT: Handbook for CSIRTs	X	X	X	X	-	-	-	X	-	X	-	-
NIST 800-61	X	X	X	X	X	-	-	X	-	X	X	X
ENISA: Gd Prt Guide for IM	X	X	X	X	-	X	X	X	X	X	-	-
ISACA: IMR	X	X	X	X	X	X	-	X	-	X	-	-
ISO/IEC 27035- 2:2016	X	X	-	X	X	-	X	X	X	X	-	X

After defining the objectives and selecting the most important topics to include in an IRP, a model of implementation was created. This model is divided into three domains that go from the essential components of a plan to the components of continuous evolution, ending with useful tools to improve the performance of its execution, thus including all topics referenced before. Figure 1 represents the proposed model.



**Figure 1:** The proposed model for implementing an IRP

### 3.1 Essential

The first domain, Essential, corresponds to the initial stage of building an IRP. The topics that are considered most important to be in an IRP are part of this step, allowing the creation of a simple but useful and functional IRP. In this phase, the topics "Purpose and Scope", "Roles and Responsibilities", "Incident Response Life Cycle", and "Simple Communication Plan" will be addressed.

#### 3.1.1 Purpose and scope

An IRP must be implemented and executed with the collaboration of key stakeholders. To understand what is reasonable to expect from the security team, it is crucial to know the community served and the services offered to the community. Furthermore, this section usually includes the mission, constituency, authority, responsibility, and services of the incident response team/process. Based on the analysed plans, some of them also place an organisation chart to locate the incident response team within the organisation (Stinson Leonard Street, LLP, 2017).

#### 3.1.2 Roles and responsibilities

The structure and responsibilities of the incident response team must be clear. An IRP must define and present each team member's roles and tasks. In addition, it should also identify the main stakeholders and their responsibilities in the incident response service, as represented in Table 3. It is essential to highlight user participation's importance since, without their input, CSIRT services' effectiveness can be significantly reduced, particularly with reporting incidents.

**Table 3:** Roles and responsibilities (example)

Title/Role	Responsibility
Chief Information Officer	- Maintain and Enforce this Procedure. (...)
Security Analyst Tier 1	- Monitor systems and activity, respond to potential security events and incidents. (...)
IT Department	- Provide IT support and expertise to the CSIRT (...)
	(...)
End Users	- Detect and report security incidents to the CSIRT (...)

Besides the presented roles, it may also include the Help Desk, Physical Security, DPO, Legal Department, Human Resources, etc.

### *3.1.3 Incident response life cycle*

The incident response life cycle is one of the key components of an IRP. It describes the organisation's step-by-step framework for identifying and responding to a security incident or threat.

Two main industry-standard (NIST and SANS) frameworks provide a few steps to the incident response life cycle. Despite their differences, there are many similarities. It is up to the organisations to choose which of them best suits their needs. It is also important to highlight that, if neither of those is considered appropriate to the organisation's reality, they can be adjusted to meet the requirements.

The IRP should focus on all phases by describing its goals, steps, and tasks. It may also be interesting to include questions, a checklist, or a flow diagram to help the reader understand the phase actions entirely.

It is common to see the available interfaces for reporting incidents or threats on the Identification or Detection phase's description since the users need to know that they should report and where to do it.

Also, independently of the chosen incident response framework, there should always be a phase where the detected event or threat is considered a security incident. When that happens, it is typically addressed by some sort of incident classification to indicate its relevance. Each team has its own way of defining severities, but it is usually assigned to a scale value (e.g., Informational, Low, Medium, and High). This triage allows the CSIRT to prioritise the incident management and even change the incident's route if considered relevant.

### *3.1.4 Simple communication plan*

When an incident response team is confronted with a potential security breach or data loss and the organisation's reputation or business is at stake, it can be overwhelming to the team because of all the technical issues related to the investigation, containment, or recovery.

The absence of a communication plan can lead to a situation where a CSIRT does not know whom to communicate with, when, and by what channel, possibly resulting in a security incident mishandling. This episode can jeopardise the organisation and possibly result in a non-compliance fine.

As a simple communication plan, the IRP should include the main stakeholders and their notification trigger. This usually comprises internal staff, regulators, law enforcement authorities, third parties, end-users/clients, among others. When possible, each stakeholder's contacts should be presented in the document or means to consult them should be provided. A simple table with the stakeholder category and its contact information should be sufficient.

## **3.2 Outlined**

This domain corresponds to the intermediate stage of constructing an IRP, aiming to complete the IRP and give it a strategy of evolution and constant improvement. In this phase, the topics "Incident and Evaluation Metrics", "Service-level Agreements", "Plan Testing and Review" and "Incident Response Training" will be covered.

### *3.2.1 Incident and evaluation metrics*

It is imperative to continuously monitor the incident handlers' tickets and interactions/actions during incident handling. The statistics collected through this monitoring provide insight into the CSIRT service's performance and the most recurring incidents, failures, or users. It is up to the organisation to define the metrics it intends to collect based on its objective. This measurement can be done manually or automatically.

### *3.2.2 Service-level Agreements (SLA)*

An SLA is a contract between a service provider and the end-user that defines the provider's service level. Therefore, this section should include the CSIRT service activation criteria, the definition of expected response times, and the indication of the service hours.

By defining service activation criteria, the CSIRT can reduce the workload and thus guarantee the incident response's success. These criteria could be, for example, to only respond to SPAM incidents if they originated

from the organisation. The activation criteria, if used, can and should be adjusted as the CSIRT grows and the responsiveness increases.

Response times are usually associated with the level of criticality provided to the incident. The time set to notify or handle a high-rated incident should be much shorter than a low-rated incident. The service's workload usually varies between 8x5, 24x7, or mixed, but these are not standard. The availability of this information allows the user to know when they will be expected to receive the handling of an incident by the CSIRT.

### 3.2.3 Plan testing and review

Periodically reviewing and updating the plan's content is necessary to update, identify gaps, and maintain its applicability at any time. Also, testing the plan affects the document's content and helps incident handlers identify poorly executed steps and aspects that need improvement.

Testing should be carried out considering a minimum periodicity (usually annual). These tests should include practical simulations of real scenarios, and the team must ensure the fulfilment of the incident response process. It is essential that the team understands how important simulations are for the service's good performance.

### 3.2.4 Incident response training

It is difficult to maintain the ability to respond to incidents effectively over time without adequate and continuous training. As in the previous section, the training of incident response teams can be carried out through practical simulations. Another way is simply through continuous training, either internally or externally. The plan should include the perspectives and objectives of the CSIRT service in this context.

For stakeholders in this process, it may be pertinent to include awareness sessions so that they know what to report and how to deal with incidents to ensure a consistent and appropriate response.

## 3.3 Support

The Support domain corresponds to the last stage of building an IRP. The implementation and use of tools such as a taxonomy or classification matrix may not be in everyone's interest, despite accelerating and improving processes. In this phase, the topics "Taxonomy", "Classification Matrix", "Playbooks", "Communication Plan", and "Useful Templates" will be covered.

### 3.3.1 Taxonomy

The incident taxonomy is a classification scheme commonly used by CSIRT to better understand and categorise the security incidents in an organisation. It also allows the organisation to have a view on the trending of incidents and threats, and to better prepare/improve the incident response team performance. Creating a taxonomy is not a simple task, and therefore it is best to use an existing classification scheme. Since the organisation will probably be exchanging some information about their security incidents with other CSIRTs or regulators, it can be useful to determine if they use a shared taxonomy<sup>1</sup>. Table 4 presents an excerpt of an example.

**Table 4:** Excerpt of an example

Category	Type	Description
Malicious Code	Infected System	System infected with malware, e.g., PC, smartphone or server infected with a rootkit. Most often, this refers to a connection to a sinkholed C2 server.
	(...)	(...)
	Malware Distribution	URI used for malware distribution, e.g., a download URL included in fake invoice malware spam or exploit-kits (on websites).
Information Gathering	Sniffing	Observing and recording of network traffic (wiretapping).
	(...)	(...)

<sup>1</sup> E.g., in Europe there is the Reference Security Incident Taxonomy Working Group (RSIT WG) scheme.

### 3.3.2 Classification matrix

As seen in the incident response life cycle, assigning a severity value allows defining a prioritisation and incident handling strategy. The classification matrix introduces a simpler and faster way to assign a level of severity to incidents, which are often assessed inappropriately or inconsistently when doing this manually. This matrix determines a default severity value according to the type of incident in question. Being a default value, it can be adjusted if considered relevant. The IRP may also include the definition of criteria for this adjustment (e.g., in the event of an incident in a critical asset, the incident's criticality should rise). Table 5 presents an excerpt from an example classification matrix.

**Table 5:** Example of a classification matrix

Category	Type	Severity by default
Intrusions	Privileged Account Compromise	S1 (High)
	Unprivileged Account Compromise	S2 (Medium)
	Application Compromise	S1 (High)
	Burglary	S2 (Medium)
Vulnerable	Vulnerable system	S4 (Informational)
	(...)	

In addition to the incident's criticality level, CSIRTs are recommended to use an information classification system in terms of sharing, the Traffic Light Protocol (TLP). Through a colour scheme (WHITE, GREEN, AMBER and RED), access to information is defined and limited, whether publicly, internally, to a limited group or only those involved.

### 3.3.3 Playbooks

Playbooks are operational procedures with practical guidelines for handling and resolving specific type/category incidents. The purpose of these documents is to guide the CSIRT and optimise and accelerate the entire process. Playbooks are often illustrated in checklists, diagrams, RACI<sup>2</sup> tables, among others. The chosen method should be the one that best serves the team members who will handle the incidents.

Usually, only a few playbooks are related to more critical incidents, including data breach, malware, or intrusions. However, the use of playbooks for all types of incidents may prove extremely useful for the team. Also, even within the same type of incident, it may still be pertinent to have different procedures according to the problem.

### 3.3.4 Communication plan

The implementation of a more detailed communication plan, in addition to the recommendations provided above, should consider three points: (i) Adapt communications to internal stakeholders and define preferred communication channels - Refers to the inclusion of entities such as management, DPO, IT department, among others. Triggers and means of contact should be defined to notify these entities, as well as what information to forward. (ii) Include practical cases that require a particular notification procedure - Situations such as data breaches may impose a different notification flow, possibly involving other entities. The definition of guidelines for these unique cases may be sorely missed in an emergency; (iii) Include notification to service providers - Providers whose service has been abused and third-party providers/services that may be useful in applying blocking measures and thus reducing the impact. It might be interesting to include some platform URLs or an indication of how to search for third party contacts (e.g., using WHOIS).

### 3.3.5 Useful templates

The primary purpose of templates is to assist and speed up some of the incident handlers/team's tasks. This section can include an incident form, a final incident report template, or even text scripts to use when there is a need to notify someone. In the latter case, templates should be prepared mainly for public communication (media statements) or third parties (abuse notification).

<sup>2</sup> Responsibility assignment matrix (Responsible, Accountable, Consulted, Informed).

#### **4. Paths to implementation**

The strategy presented includes a phased implementation model, and the application of this model may differ from entity to entity. There is no single path. Each organisation is different, and, therefore, it is necessary to assess its objectives, needs, and capabilities concerning the management of security incidents. Thus, the plan to implement should reflect the organisation's maturity in order not to be considered inadequate or have a careless or rare use. The model phases are executed sequentially, starting in the Essential domain, continuing to Outlined and ending in Support. The plan was made so that the entities can make their own path. An organisation can choose only to make a simple plan, namely in situations where investment in the incident response service is reduced or made in an ad-hoc approach by IT employees. On the other hand, an organisation that considers itself more mature may choose to perform the first two phases, either sequentially or at once. In the case of organisations already mature and consolidated, they will probably try to implement a complete IRP, going through all this model's domains.

The presented model's main utility is that, regardless of the phase in which an organisation is, it has a useful and ready-to-use plan, reducing the waiting time to start the services. Finally, although the Support domain corresponds to the third phase, it is important to mention that it can be started at any time after the first domain's consolidation. The implementation of the second phase is not mandatory, although recommended. It is also important to remember that, to achieve the plan's success, participation and support from the administration is compulsory, as well as the definition of measures to evaluate the service's success.

#### **5. Conclusions and future work**

As security threats and incidents rise, it surges a need for investing in monitoring and protecting information systems. In case of an incident, the response process must be assertive and quick to mitigate the impact and reduce costs to the organisation.

The development and use of a complete, practical, and straightforward IRP become essential for responding to security incidents. However, the implementation of such a document requires time, knowledge, and some effort to analyse the norms and frameworks. Even so, these standards, despite presenting the main ideas and including some tips, do not provide an implementation methodology.

Through the research and study of the documents of reference, related articles, and incident response plans of several organisations, it was developed a practical strategy for implementing an IRP that can be used to complement the existing standards and frameworks.

After consolidating the most important and referenced topics in the analysed documentation, a model was developed, divided into three domains: Essential, Outlined and Support. Each domain includes practical tips and examples of what and how to integrate each subject.

The adoption of the proposed model allows answering several essential questions in the process of creating an IRP that is effectively used and put into practice. These include questions such as "What is a security incident for the organisation?"; "Who are the stakeholders?"; "What is the response procedure?"; "To who communicate in a data breach incident?"; "Which of the incidents to give priority to?"; among others. Subjects that promote evaluation and progressive improvement for the process and team are also included.

The proposed strategy for implementing this plan may differ from organisation to organisation, as the model was designed to be adjustable to each one's reality. A plan with only the first domain of the consolidated model can already be considered a useful and applicable plan in several realities. If considered insufficient, the remaining domains should be explored. Through this strategy, the implementation of an IRP is phased and modular, allowing it to evolve according to organisations' needs.

The model and strategy presented in this work still need to be validated, at least in applying it to an organisation to measure its effectiveness. In this regard, it is recommended that future work will focus on the applicability of this model to different organisations and across multiple verticals, e.g., government, industry, and academia. Since this strategy or model may also not be adequate when using it against an organisation that outsources the security incident management services, it should be reviewed to increase the applicability coverage.



## Acknowledgements

This work was supported by Portuguese national funds through the FCT - Foundation for Science and Technology, I.P., under the project UID/CEC/04524/2020.

## References

- Al-Dhaqm, A., Razak, S. A., Siddique, K., Ikuesan, R. A., & KEBANDE, V. R. (2020). Towards the Development of an Integrated Incident Response Model for Database Forensic Investigation Field. *IEEE Access*, 8, 145018-145032. doi:10.1109/ACCESS.2020.3008696
- Brownlee, N., & Guttman, E. (1998). *RFC2350: Expectations for Computer Security Incident Response*. USA: RFC Editor.
- Criminal Justice Information Center. (2019). *Example Incident Response Plan*. Retrieved January 13, 2021, from State of Michigan: [https://www.michigan.gov/documents/msp/Example\\_Incident\\_Response\\_Policy\\_666657\\_7.pdf](https://www.michigan.gov/documents/msp/Example_Incident_Response_Policy_666657_7.pdf)
- Empresa Brasileira de Serviços Hospitalares. (2018, July). *PGI – Plano de Gerenciamento de Incidentes do HU-UFJF*. Retrieved January 13, 2021, from Empresa Brasileira de Serviços Hospitalares - EBSERH: [http://www2.ebserh.gov.br/documents/222346/866032/Plano+ Gerenciamento Incidentes ajustes JUL 18.pdf/2e6d868a-39fa-4273-a889-7bcb2a3f3867](http://www2.ebserh.gov.br/documents/222346/866032/Plano+Gerenciamento+Incidentes+ajustes+JUL+18.pdf/2e6d868a-39fa-4273-a889-7bcb2a3f3867)
- ENISA. (2010). *Good Practice Guide for Incident Management*. Publications Office.
- International Organization for Standardization. (2016). *ISO/IEC 27035-2:2016: Information Security Incident Management*.
- ISACA. (2012). *Incident Management and Response*.
- IT Governance Institute. (2007). *CobIT 4.1: Framework, Control Objectives, Management Guidelines, Maturity Models*. Rolling Meadows: IT Governance Institute.
- JSM Marketing Services. (2017, November). *Incident Response Plan*. Retrieved January 13, 2021, from JSM Marketing: [https://jsm-marketing.com/wp-content/uploads/2018/03/JSM\\_Incident-Response-Plan-2018.pdf](https://jsm-marketing.com/wp-content/uploads/2018/03/JSM_Incident-Response-Plan-2018.pdf)
- Mooi, R., & Botha, R. A. (2016). Context for the SA NREN Computer Security Incident Response Team. *2016 IST-Africa Week Conference*, (pp. 1-9). doi:10.1109/ISTAFRICA.2016.7530662
- Onwubiko, C., & Ouazzane, K. (2020). SOTER: A Playbook for Cybersecurity Incident Management. *IEEE Transactions on Engineering Management*, 1-21. doi:10.1109/TEM.2020.2979832
- OSCIO Enterprise Information Services. (2016). *Information Security Incident Response Plan (SAMPLE)*. Retrieved January 13, 2021, from State of Oregon: <https://www.oregon.gov/das/oscio/documents/incidentresponseplantemplate.pdf>
- Pessoa, João; EGPr-TIC. (2016). *Processo de Gerenciamento de Incidentes*. Retrieved January 13, 2021, from Tribunal Regional do Trabalho 13ª Região - Paraíba: <https://www.trt13.jus.br/institucional/governanca/publicacoes/trt-13/setic/escritorio-de-processos/processo-de-gerenciamento-do-incidentes/Processo%20de%20Gerenciamento%20de%20Incidentes.pdf>
- PROCEMPA. (2020, August). *Plano de Resposta a Incidentes de Segurança e Privacidade*. Retrieved January 13, 2021, from Prefeitura de Porto Alegre: [https://prefeitura.poa.br/sites/default/files/usu\\_doc/sites/procempa/Plano%20de%20Resposta%20a%20Incidentes.pdf](https://prefeitura.poa.br/sites/default/files/usu_doc/sites/procempa/Plano%20de%20Resposta%20a%20Incidentes.pdf)
- Proffitt, Timothy; SANS. (2007). *Creating and Managing an Incident Response Team for a Large Company*. *SANS.edu Graduate Student Research*.
- Scarfone, K. A., Grance, T., & Masone, K. (2012). *SP 800-61 Rev. 2. Computer Security Incident Handling Guide*. Gaithersburg, MD, USA: National Institute of Standards & Technology.
- Stinson Leonard Street, LLP. (2017, October). *Security Incident Response Plan [Sample]*. Retrieved January 13, 2021, from Carolinas Credit Union League: <https://www.carolinasleague.org/resource/collection/B728697E-626E-4B25-8C68-28A43CF6536F/3%20B%20Pirnie%20Handout%20Sample%20Incident%20Response%20PI.pdf>
- Texas Southmost College. (2018). *IT Security Incident Response*. Retrieved January 13, 2021, from Texas Southmost College: [http://archive.tsc.edu/images/helpdesk/TSC\\_-\\_IT\\_3.6\\_IT\\_Security\\_Incident\\_Response.pdf](http://archive.tsc.edu/images/helpdesk/TSC_-_IT_3.6_IT_Security_Incident_Response.pdf)
- Tulane University. (2020). *Computer Incident Response Plan*. Retrieved January 13, 2021, from Tulane University | Information Technology: <https://it.tulane.edu/computer-incident-response-plan>
- Universidade Federal do Rio de Janeiro – UFRJ. (2018). *Plano de Gestão de Incidentes de Segurança da Informação*. Retrieved January 13, 2021, from Diretoria de Segurança da Informação e Governança - UFRJ: <https://www.security.ufrj.br/wp-content/uploads/2019/02/Plano-de-gest%3a3o-de-incidentes-SegTIC.pdf>
- University Carnegie Mellon. (2020, September). *Computer Security Incident Response Plan*. Retrieved January 13, 2021, from University Carnegie Mellon: <https://www.cmu.edu/iso/governance/procedures/IRPlan.html>
- University of Adelaide. (2018). *Data Breach Response Plan*. Retrieved January 13, 2021, from The University of Adelaide: <https://www.adelaide.edu.au/policies/62/?dsn=policy.document;field=data;id=8225;m=view>
- University of Alabama. (2016). *IT Practice Procedure Security Incident Response Plan*. Retrieved December 18, 2020, from The University of Alabama: <https://oit.ua.edu/services/security/report/>
- West-brown, M., Stikvoort, D., & Kossakowski, K.-P. (1999). *Handbook for Computer Security Incident Response Teams (CSIRTs)*.
- Western Oregon University. (2009). *Data Security Breach Incident Response Plan*. Retrieved January 13, 2021, from Western Oregon University: [https://wou.edu/ucs/files/2015/11/WOU\\_Incident\\_Resp\\_Plan.pdf](https://wou.edu/ucs/files/2015/11/WOU_Incident_Resp_Plan.pdf)

# Are Encrypted Protocols Really a Guarantee of Privacy?

Jan Fesl<sup>1,2</sup>, Michal Konopa<sup>1</sup>, Jiří Jelínek<sup>1</sup>, Yelena Trofimova<sup>2</sup>, Jan Janeček<sup>1,2</sup>, Marie Feslová<sup>1</sup>, Viktor Černý<sup>2</sup> and Ivo Bukovsky<sup>1</sup>

<sup>1</sup>University of South Bohemia, Branisovska 31, Ceske Budejovice, Czech Republic

<sup>2</sup>Czech Technical University in Prague, Faculty of Information Technology, Thákurova 9, Prague, Czech Republic

[jfesl@prf.jcu.cz](mailto:jfesl@prf.jcu.cz)

[konopm05@prf.jcu.cz](mailto:konopm05@prf.jcu.cz)

[jjelinek@prf.jcu.cz](mailto:jjelinek@prf.jcu.cz)

[trofiyel@fit.cvut.cz](mailto:trofiyel@fit.cvut.cz)

[janecek@fit.cvut.cz](mailto:janecek@fit.cvut.cz)

[dolezm01@prf.jcu.cz](mailto:dolezm01@prf.jcu.cz)

[cernyvi2@fit.cvut.cz](mailto:cernyvi2@fit.cvut.cz)

[ibuk@prf.jcu.cz](mailto:ibuk@prf.jcu.cz)

DOI: 10.34190/EWS.21.047

**Abstract:** Most internet traffic is being encrypted by application protocols that should guarantee users' privacy and anonymity of data during the transmission. Our team has developed a unique system that can create a specific pattern of traffic and further analyze it by using machine learning methods. We investigated the possibility of identifying the network video streams encrypted within the HTTPS protocol and explored that it is possible to identify a particular content with a certain probability. Our paper provides a methodology and results retrieved from the real measurements. As the testing data set, we used the streams coming from the popular platform Youtube. Our results confirm that it is possible to identify encrypted video streams via their specific traffic imprints, although it should not be possible due to the used encryption.

**Keywords:** internet traffic, encrypted video stream, identification, data traffic pattern, machine learning

---

## 1. Introduction

In the era of many cyber attacks, user anonymity represents a very important aspect of Internet functionality. Many services like Internet-banking, electronic email due to the potential risk require the encrypted communication between the client and server-side. The popular platforms like Youtube or Netflix adopted a related encryption mechanism for video network streaming - all content is currently available via HTTPS protocol with the TLS 1.3 encryption. The enabled encryption should offer to the end-users the feeling of anonymity - nobody should be able to distinguish the content of the encrypted video stream. The traffic between the end-user and the streaming server could be easily captured by network probes and subsequently analyzed. The network video-stream has a unique property that periodically repeats - the users periodically look at the same movies. The detection of the user watched content may act inappropriately, but there exists i.e. malicious traffic like child-pornography and its detection has substantial meaning. The motivation for our research was to verify the idea if, for a periodically transmitted network stream, it is possible to create a specific fingerprint allowing the stream identification.

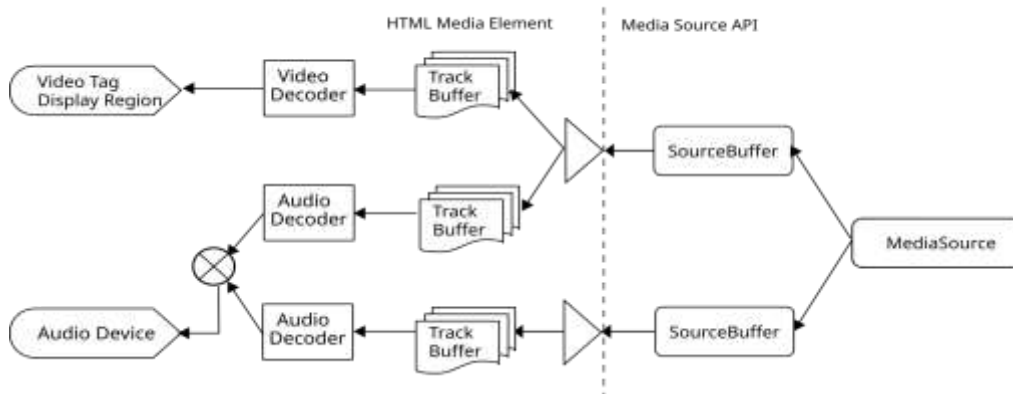
Our research group developed a unique solution based on Apache big-data ecosystem which allows us to capture and efficiently visualise and analyse the network streams - data packed in IP datagrams. Based on the obtained preliminary results, we proposed three different strategies confirming our theory about the possible network stream pattern.

## 2. Related background

### 2.1 Network video streams and their transmission

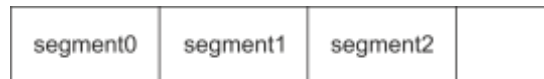
For a long time browsers needed to use some external player for playing videos. HTML5 format was created with that fact in mind and supports streaming of some video source directly in the browser if the last one supports the corresponding codecs. Together with W3C specification called Media Source Extensions (MSE) it allows for complex use cases such as changing video quality, multiple languages and others. As of 2020, HTML5 video is the only widely supported video playback technology in modern browsers. MSE extends HTML5 (HTMLMediaElement) to allow JavaScript to generate media streams for playback. It defines a MediaSource

object which has one or more SourceBuffer elements. For example, video and audio data can be separated into two source buffers (Fig. 1).



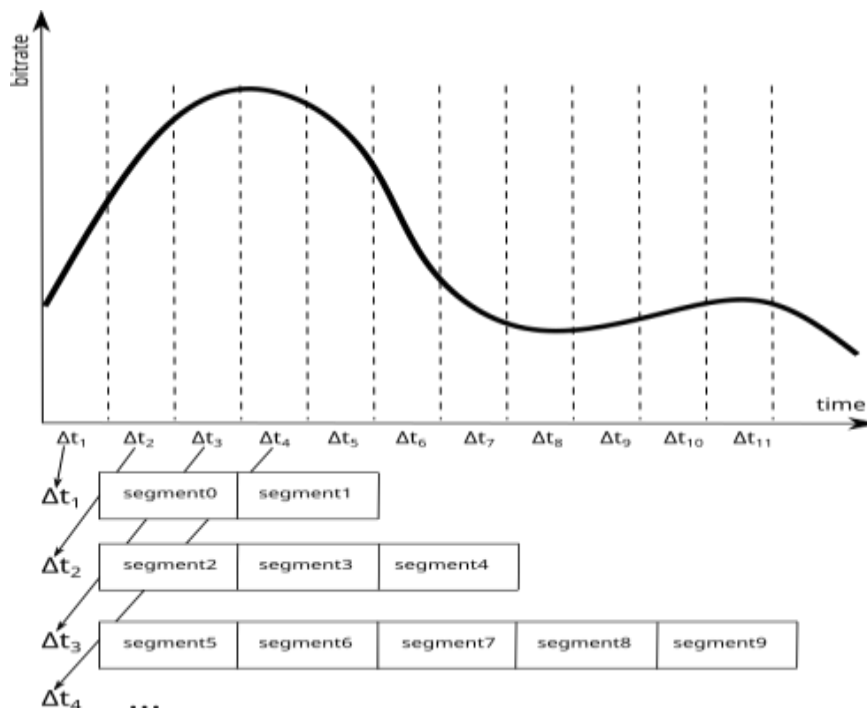
**Figure 1:** Implementation of sending byte streams to media codecs within web browsers. This principle has universal validity and it is used by all usually available browsers.

Furthermore, audio/video content is split into segments. So, when the first one is obtained, the video can start playing.



**Figure 2:** Visualisation of segments in a media file. Each segment is received by the javascript player and stored into the playing buffer

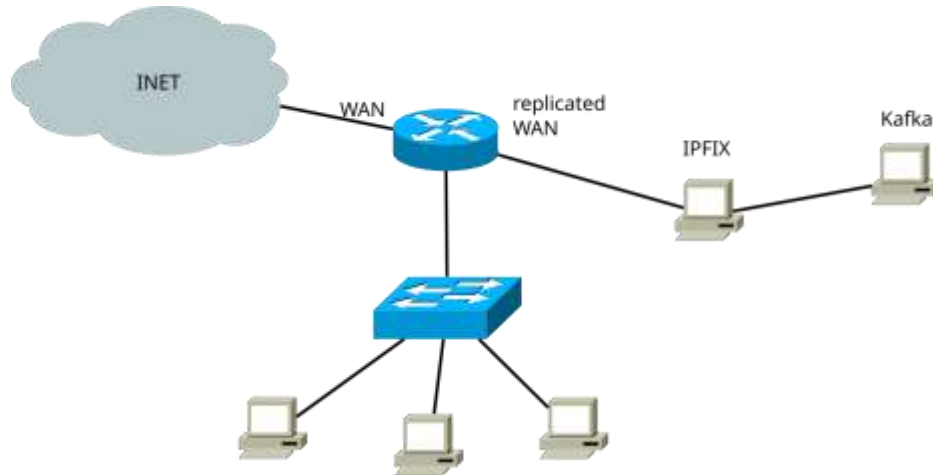
Various scenes in the video may require different amounts of data to be sent, thus distinct numbers of segments are required per the same amount of time  $\Delta t$ , depicted in Fig. 3. This creates a unique fingerprint for each video.



**Figure 3:** The segment numbers dependency on the bitrate. This principle demonstrates how the bitrate can change depending on the specific features of video scenes

## 2.2 Capturing of network traffic and generating of fingerprints

The measurement scenario was that all traffic between the client's web browser was forwarded through the network router which mirrored the specific traffic to the network traffic probe. The complete scenario can be seen in Fig. 4. The central router mirrors entire traffic to the network probe that makes a preliminary classification and reports the summary information via NetFlow/IPFIX protocol to the messaging system (Apache Kafka) serving as a pipeline. All records stored in the pipeline are subsequently preprocessed or visualised and further analyzed by detection modules.



**Figure 4:** The network infrastructure serving for the creation of data-flows fingerprints. Such a solution is efficient and very easily scalable

## 2.3 The current state of the art in encrypted streams detection

Recently, the possibility of identifying encrypted streams via time traffic behaviour was proposed by Hejun and Liehuang (2019) via dynamic time warping (DTW) with fundamental clustering. Thus, it is apparent that time analysis methods such as DTW and potentially multiscale DTW methods, e.g., Dilmi (2020) et al., have a certain potential for encrypted stream identification. So, there is partial inspiration by recent studies and principally relevant concept of joint time-frequency analysis, e.g., Chen and Hao (1999). The adaptive streaming classification using some machine learning methods was proposed by Dubin (2017), Li (2018) used heuristics for Youtube video identification and Reed (2017) tested another heuristics for Netflix videos detection. Convolutional Neural Networks were used for encrypted video classification by Schuster (2017) and SVM classifiers were applied by Shi (2016). In Shi (2021) the authors used NLP for the video source identification. Wu introduced the differential video fingerprints in Wu (2020) and fuzzy logic for encrypted stream identification was applied by Zu (2016).

Thus we propose and experimentally demonstrate that the time distribution function of transmitted data has the potential to be a 1-D specific imprint of encrypted video streams. Furthermore, we propose a novel time-spatial distribution-specific 2-D imprint that simultaneously characterizes both time behaviour and spatial distribution of transmitted data volumes, which results in a 2-D image that visually demonstrates the meaning of the characteristic imprint of the whole transmitted (encrypted) data stream. Such time-spatial representation, i.e., the imprint, is then visually comprehensible to humans, and it is also suitable for further image processing or processing by machine learning methods.

Further in the text, the parts of transmitted data within an empirically chosen short period are denoted briefly as (data) chunks.

## 3. Applied methods

This section proposes several concepts for obtaining features (further called imprints) and discusses their processing techniques for encrypted video stream identification.

The first proposed concept is based on merging representations of time volume distribution and spatial volume distribution of streaming data in subsection 3.1; this imprint is also suitable for visual understanding. Thus, it has potential for image processing, including machine learning techniques.

The next concept is based on an analysis of arrival times of individual packets. It is based on the assumption that the amount of transmitted data in the video file tends to change dynamically, with respect to, for example, passages of dynamic scenes, which should also be reflected in the sequence of time intervals in which the client-side buffer is filled. Each such sequence can then be transformed into various forms, such as a graph, to create an imprint of a particular video.

### 3.1 Distribution 2-D imprint in time domain and spatial domain

First, we propose that the time distribution function of transmitted volumes (so-called chunks) of encrypted video streams may serve as one basic part of the imprint for identification of encrypted video streams.

From the transmitted packet size evidence for a given stream, the chunk volumes are calculated as follows

$$x(t) = \sum_{v t_{\kappa}} v(t_{\kappa}) \text{ [byte]}; (t - \Delta t) < t_{\kappa} < t, \quad (1)$$

where  $x(t)$  is the data chunk volume,  $v(t_{\kappa})$  is the volume of all packets that were downloaded at time  $t_{\kappa}$  within the actual time interval  $(t - \Delta t, t >)$ , and  $\Delta t$  is an empirically selected granularity parameter, i.e., a short constant time period (usually <500msec), and  $t$  is the continuous time index in general.

Then, one 1-D imprint of the encrypted video stream may be proposed as a conventional distribution function in time domain as follows

$$F_x(t) = \sum_{\tau=0}^t x(\tau) \text{ [byte]}; t \in \langle 0, T \rangle \quad (2)$$

where  $F_x(t)$  denotes the chunk volume distribution function in the time domain, and  $T$  is the total data transmission (snapshot) time.

However, we propose to extend the imprint also with the spatial volume distribution of data traffic. Thus utilizing the above notation, the 2-D imprint that involves time-volume distribution as well as spatial-volume distribution is introduced as follows. Alongside with the time distribution function  $F_x(t)$  as in (2), the information about the spatial distribution of data chunk (1) volumes can be utilized to enhance the feature, i.e., to enhance the imprint of encrypted streams for their better classification. Given the data chunk volumes of each encrypted video stream as in (1), we can utilize the spatial-volume distribution function, of chunks, i.e., in volume domain, as follows

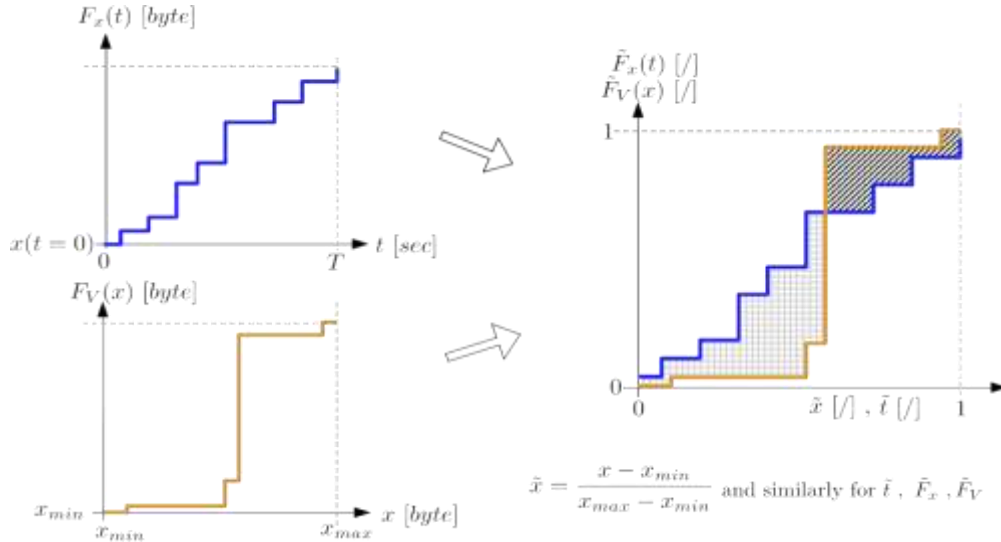
$$F_V(x) = \sum_{\xi=x_{min}}^x \xi \text{ [byte]}; x_{min} \leq x \leq x_{max}, \quad (3)$$

where  $F_V$  denotes the chunk volume distribution function in volume domain (spatial domain),  $x$  is the chunk size as in (1), and  $x_{min}, x_{max}$  is the minimum and maximum chunk size respectively.

Then, the 2-D imprint is obtained by merging the time-volume distribution  $F_x(t)$  and  $F_V(x)$  spatial-volume distribution of data chunks by drawing them within a single image after MinMax normalization of all variables as in Fig. 5. In experimental subsection 4.1, the 2-D imprints are shown for the random set of encrypted video data streams with further discussion.

### 3.2 Classification using start times of the chunks

The experiment aims to identify the playback of a specific video on a specific playback software by monitoring data packets. The video is downloaded using the player's memory buffers. These are used to compensate for fluctuations in the data stream and ensure smooth video playback. The filling of these buffers proceeds in chunks identified in the packet stream. The timing of these chunks then depends on the need for data from the buffer. This depends on the video bitrate. If this flow is variable, then the more complex the description of the scene in the video, the more data is needed to describe it, and the more frequent the buffer filling. If we accept the assumption that the buffer is filled in each chunk, it is then possible to work with chunk start times as with discrimination information. From the intervals between them, we can get a unique imprint of the video.



**Figure 5:** The principal sketch of 2-D time-volume and spatial-volume distribution imprint as one of the features for encrypted video stream identification; the 2-D imprint with MinMax normalized variables is on the left

A two-step procedure using a classification model inspired by the ART2, developed by Carpenter & Grossberg, (1987), the paradigm was used to solve the given task. The model contains two layers of cells, with the first (input), used to present input patterns and the second (output) containing one cell for each output category. The model activity is based on finding similarities between the input patterns using the Hamming distance. A model is a clustering tool that uses a tolerance parameter to control the number of clusters that will be identified. The larger the tolerance, the more input patterns classified into one cluster. On the contrary, low tolerance will ensure a larger number of them.

### 3.3 Classification based on packet arrival times

This method is based on a similar idea to that described in paragraph 3.2. The amount of transmitted data in the video file tends to change dynamically - concerning, for example, passages of dynamic scenes. From a logical point of view, this should also be reflected in the sequence of time intervals in which the buffer on the client-side is filled. According to our measurements, data are downloaded to the buffer at very narrow time intervals (typically in the order of several hundred milliseconds) followed by longer pauses, when no data are transmitted. By plotting the rank number of each incoming packet on the X-axis and displaying the time of its real arrival - relative to the time of arrival of the first video packet - on the Y-axis, we get a graph with a "stepped" course. The length of the "step" indicates how many packets were transmitted within the respective time interval, the height of the "step" then indicates the length of the pause until the beginning of the subsequent data transmission. The overall shape of the graph constitutes the imprint of a specific video.

Mathematically expressed:

$$F_{packettime}: D_{packetrank} \rightarrow \langle 0, T \rangle, \text{ where:} \quad (4)$$

$$D_{packetrank} = \{1, 2, \dots, N\} \quad (5)$$

N is the total number of transmitted video packets, T is the time of the last transmitted packet.

## 4. Practical measurements and results

Our testing data set was created as follows. We randomly selected 12 different video streams from the Youtube platform and measured for every 2 verified repetitions. All selected videos were approx. the same duration. The measurement conditions i.e. used web browser, operating system, user account, etc. were during the experiments fixed. For each record was performed the filtration of deduplicated packets and assembled the network flow like a time consequence of packets. On such network flows, there were applied the methods in detail described in sections 4.1-4.4. For each network stream, there exist two main factors for classification - the total volume of the received data and the duration of the transmission, both parameters were consistent for all

measured items. The volume of the data and transmitting time determine the video characteristics i.e. resolution, quality (used encoding codec) or changings between video frames. We assumed that all network streams were played entirely without stop or replay.

#### 4.1 Distribution 2-D imprint in time domain and spatial domain

According to subsection 3.1 and Fig. 5, the 2-D imprint actually consists of three features, the time-volume distribution, spatial-volume distribution, and the total size of the evaluated stream snapshot. In Fig.6 and Fig. 7, the original video file is denoted with a number, and the instance of streaming is denoted by a small letter, e.g., “file 1c” denotes the 3rd stream instance of video file 1.

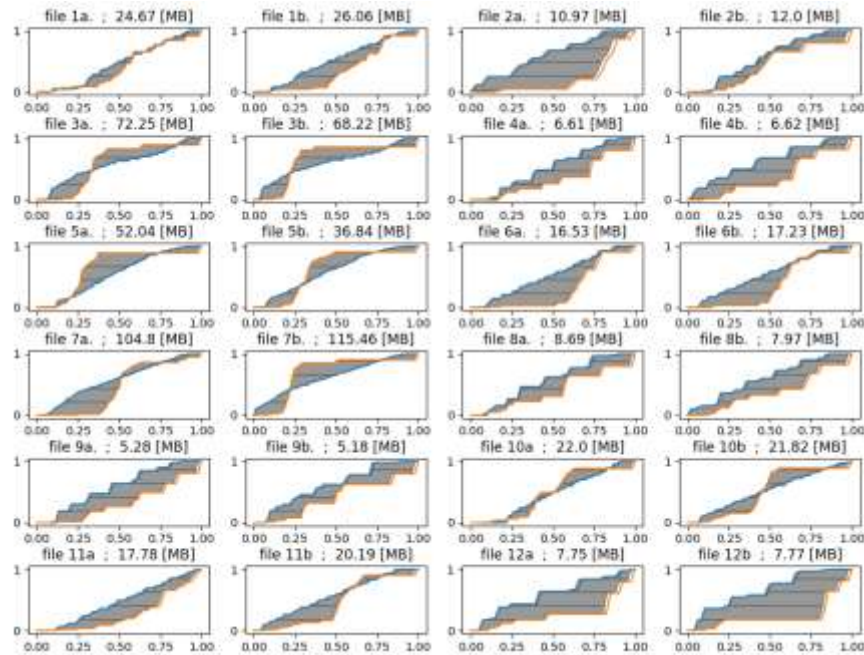


Figure 6: The visualisation of the 2D imprints of measured video streams for a specific time window and granularity  $\Delta t = 2$  [sec]

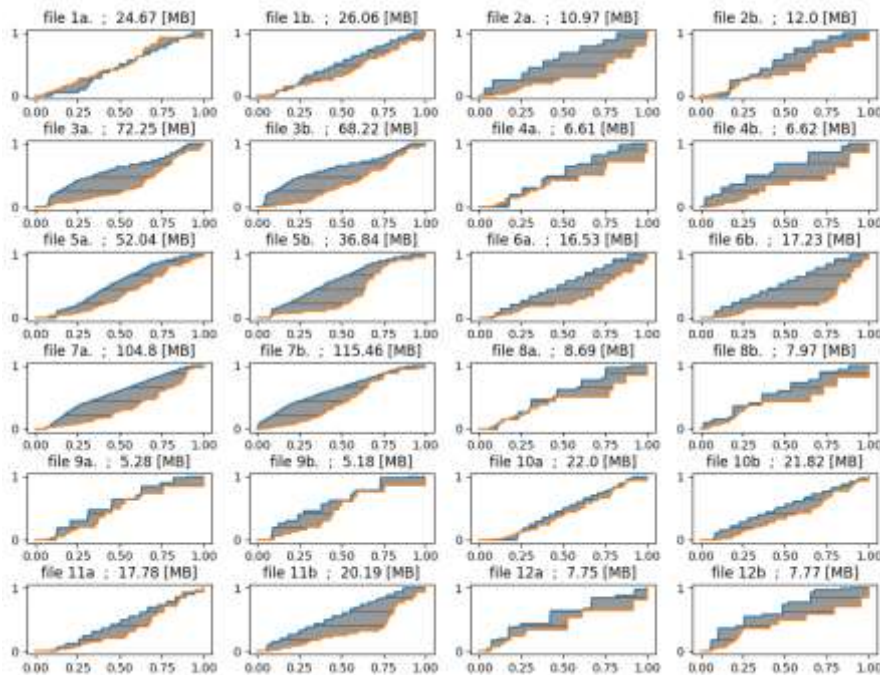


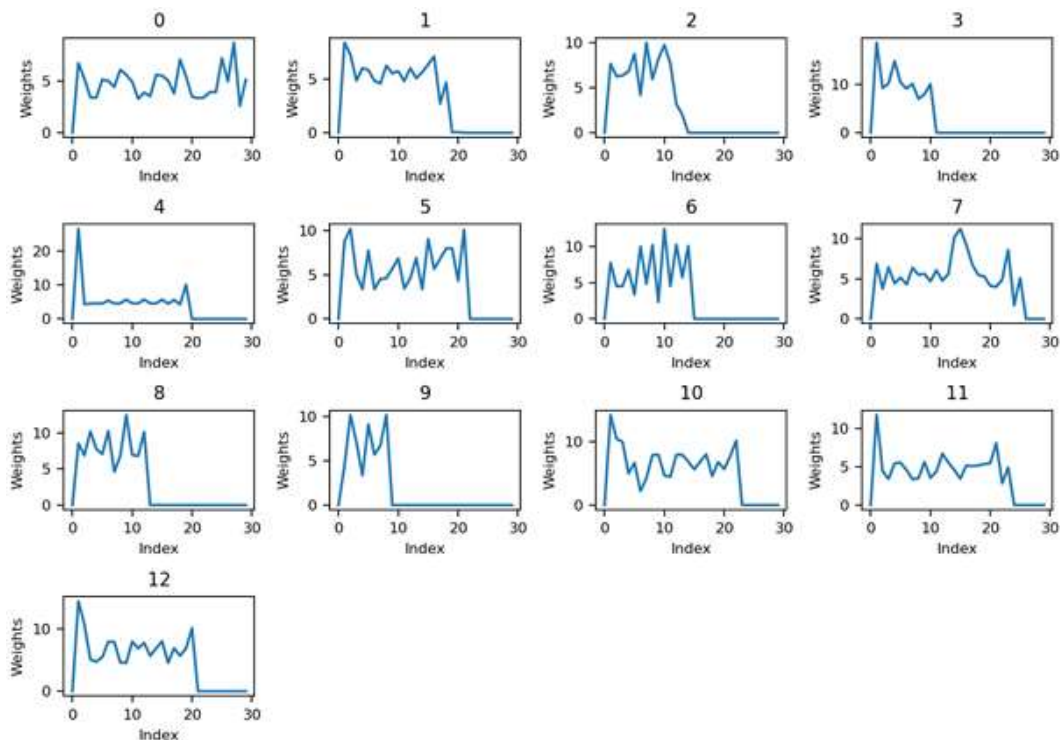
Figure 7: The visualisation of the 2D imprints of measured video streams for a specific time window and granularity  $\Delta t = 0.1$  [sec].



The results in Fig. 6 and Fig. 7 are intended mainly for visual presentation of the method and its results to kind readers in this paper. Being shown always on two streaming instances for 12 different encrypted video files, we can see that the 2-D imprints can be used to match the same encrypted video files and they can distinguish between different sources (please notice that similarity of file size is a feature to be considered first). It shall be highlighted that the granularity parameter  $\Delta t$  given in (1) plays its crucial role during the stream identification process, and the similarity of the imprints varies with decreasing  $\Delta t$ , which complies with concepts observed the related multiscale techniques, such the multiscale DTW etc. For example, the encrypted stream 2a for large granularity  $\Delta t = 2 [sec]$  in Fig. 7 might be confused with the imprints of file 12 (though the file size feature should avoid that); however, the imprints 2a and 12a&b are then very much distinguished for granularity  $\Delta t = 0.1 [sec]$ . Thus we propose the developed method in 3.1 (and demonstrated here in 4.1) to be a promising part of the identification system for encrypted video imprints that consist of this method and further proposed methods in this paper.

#### 4.2 Classification using start times of the chunks

In the first level of the process, the data streams were analyzed, describing the playback of the set of videos on a specific player. The aim was to identify individual chunks of buffer filling. Every chunk was classified into a separate category. The chunk's start times were determined and the differences between every chunk and the following one were calculated. Thus, this process transformed the input data on packet transmission into a vector with a length corresponding to the number of chunks minus one. The experimental data set was obtained by playing a set of videos, each played at least twice. The data obtained from the first level were adjusted to the same length by adding zeros and then used in the above classification model on the second level. It was experimentally found that the data from the first step contain discriminatory information. It enables the classification and thus identification of the playback of a particular video. The centroids of the individual clusters are shown in the following figure.



**Figure 8:** Visualization of centroids representing the output classification categories in the second step of processing

The input videos have been classified into the categories depicted in Figure 9. The classification algorithm based on non-supervised neural networks is able to well distinguish different patterns. The performance (some inaccuracies) could be probably improved by the usage of more measurement examples in the data set.

When identifying the particular video, the described procedure should be combined with other techniques, such as classification based on the total volume of data, or other techniques mentioned in the article.



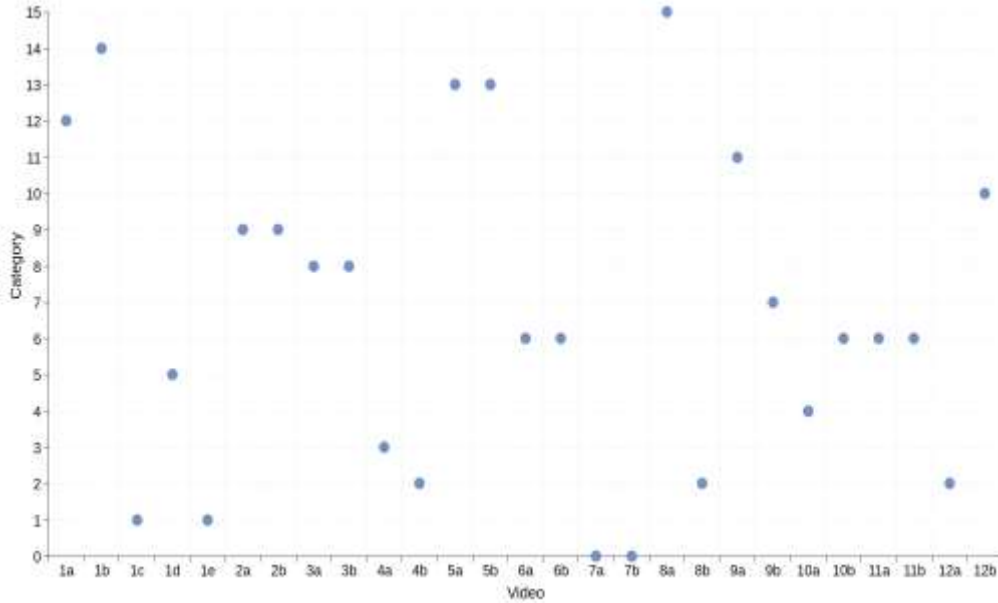


Figure 9: The classification of video streams into categories created by the neural network ART2 and post-processing

### 4.3 Classification based on packet arrival times

As part of the analysis, we created graphs for all versions of all tested videos showing the dependence of the time of the incoming packet on the packet rank - see Fig. 10. For each video, the graphs of its individual versions were compared. Each pair of graphs was synchronously compared from the packets with the highest packet rank to the beginnings of both sequences. This was due to the fact that we were unable to satisfactorily filter out other data (often ads, trailers) that preceded the test video's own data. We mainly compared the relative heights and lengths of the corresponding steps of both courses, as well as the overall shape of the graphs. In 7 videos out of a total of 12, there was a decent match in all versions of the concrete video, in 3 videos the match of the versions was significantly lower and in 2 there was a relatively large difference in all measured versions. The lowest match rate was recorded for videos with a small amount of data, typically up to 10,000 packets. These videos in the graph showed, compared to the larger ones (even over 40,000 packets), a smaller number of steps, i.e. fewer time intervals in which the data came.

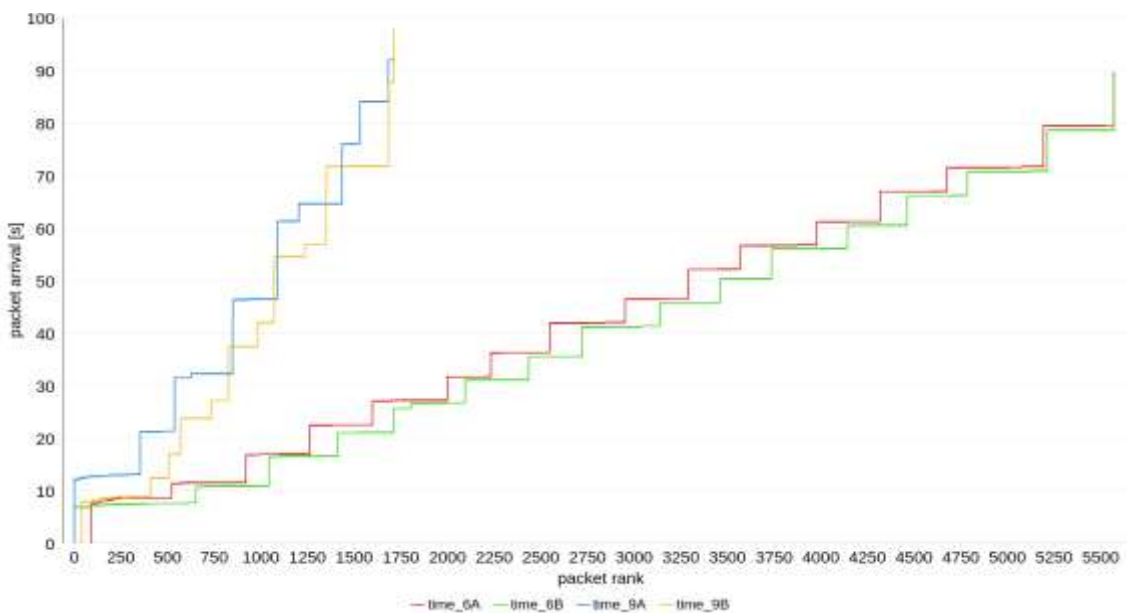


Figure 10: The best (file 9) and worst(file 6) matches in the group of tested videos

## 5. Conclusion

In our paper, we introduced the topic regarding encrypted video stream transmissions. We described the principle of packet transmission used in video streams which are played by javascript players in web browsers and highlighted the possibility of retrieving the stream patterns. For the verification of correctness of our idea, we created a testing data set containing real streams from Youtube platform. We described the main principles contained in these methods and applied them to our data set. All methods were able to classify the video streams with the good ability which confirmed our idea that it is possible to identify the encrypted video streams.

The streaming of malicious content is really a big problem of the current Internet, the consequences of this fact can be at many levels i.e. the moral and financial. The proposed technique can be used for online stream identification and efficient stream blocking. Further development will be targeted on the robust stream detection, identification of parameters allowing exact stream naming and selection of the stream description features.

The encrypted video-streams are a very good example of the fact that although the content is safely encrypted, it is possible to create a specific pattern allowing the exact stream identification. The video streams do not guarantee privacy due to the specified reasons. This fact can well help to distinguish and eliminate malicious content.

## Acknowledgements

The authors would like to acknowledge the University of South Bohemia in České Budějovice and Czech Technical University in Prague for providing the background necessary for the paper creation. Further, the authors would like to acknowledge the association Cesnet for providing the financial support used for the acquisition of technical accessories.

## References

- Chen, V., C., Ling, H. (1999) "Joint time-frequency analysis for radar signal and image processing", *IEEE Signal Processing Magazine* 16, pp 81–93.
- Dilmi, M.D., Barthès, L., Mallet, C., Chazottes, A.(2019) "Iterative multiscale dynamic time warping (IMs-DTW): tool for rainfall time series comparison", *International Journal of Data Science and Analytics* 10, pp 65–79.
- Dubin, R., Dvir, A., Pele, O., Hadar, O. (2017) "I Know What You Saw Last Minute—Encrypted HTTP Adaptive Video Streaming Title Classification", *In IEEE Transactions on Information Forensics and Security*, December, Vol. 12, No. 12, pp 3039-3049.
- Li, F., Chung, J., Claypool, M. (2018) "Silhouette: Identifying YouTube Video Flows from Encrypted Traffic", *Proceedings of the 28th ACM SIGMM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV '18)*, Association for Computing Machinery, New York, NY, USA, pp 19–24.
- Reed, A., Kranch, M. (2017) "Identifying HTTPS-Protected Netflix Videos in Real-Time", *In Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy (CODASPY '17)*, Association for Computing Machinery, New York, NY, USA, pp 361–368.
- Schuster, R., Shmatikov, V., Tromer, E. (2017) "Beauty and the burst: remote identification of encrypted video streams", *In Proceedings of the 26th USENIX Conference on Security Symposium (SEC'17)*, USENIX Association, USA, pp 1357–1374.
- Shi, Y., Biswas, S. (2016) "Protocol-independent identification of encrypted video traffic sources using traffic analysis," *2016 IEEE International Conference on Communications (ICC)*, Kuala Lumpur, pp 1-6.
- Shi, Y., Feng, D., Cheng, Y., Biswas, S. (2021) "A natural language-inspired multilabel video streaming source identification method based on deep neural networks", *SIVIP* (2021).
- Wu, H., Yu, Z., Cheng, G., Guo, S. (2020) "Identification of Encrypted Video Streaming Based on Differential Fingerprints", *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Toronto, ON, Canada, pp 74-79.
- Zhang, H., Hou, R., Lei, Yi., Meng, J., Pan, Z., Zhou, Y. (2016) "Encrypted data stream identification using randomness sparse representation and fuzzy Gaussian mixture model", *Proc. SPIE 10011*, First International Workshop on Pattern Recognition, July.

# Targeting in All-Domain Operations: Choosing Between Cyber and Kinetic Action

Tim Grant<sup>1</sup> and Harry Kantola<sup>2</sup>

<sup>1</sup>R-BAR, Benschop, The Netherlands

<sup>2</sup>Finnish National Defence University, Helsinki, Finland

[tim.grant.work@gmail.com](mailto:tim.grant.work@gmail.com)

[harry.kantola@mil.fi](mailto:harry.kantola@mil.fi)

DOI: 10.34190/EWS.21.045

**Abstract:** Targeting is the process of selecting and prioritizing targets and matching the appropriate response to them. The process is the same whether the response is cyber, kinetic, or some combination, but there are major differences between cyberspace and the physical (i.e. land, sea, air, and space) domains. Cyber operations are often stealthy and can proceed much faster than kinetic operations. Cyber targets differ greatly from their physical counterparts, and can – if desired – be engaged with no permanent damage. Cyber effects can more easily cross geographical boundaries and jurisdictions, but with an increased risk of collateral damage. Western militaries are transitioning from joint operations to all-domain (or multi-domain) operations. Until now, targeting has been done separately by domain. In future, units will have capabilities in all five domains, constructing cross-domain kill-chains from sensor to shooter. A new targeting choice will then arise: do we engage this target in cyberspace or through a physical domain? Intuitively, it would seem that kinetic action is better suited to destroying assets and denying access to an area, while cyber action lends itself to deception. However, the combination of cyber and kinetic action may be better still. One example is Operation Orchard in which the Israeli Air Force bombed a suspected nuclear reactor at Al Kibar, Syria, in 2007. To hide the ongoing air raid, the Israelis took over the Syrian air defence system using network attack techniques, feeding it a false sky picture. This paper proposes an approach for choosing between cyber and/or kinetic action as part of the targeting process in all-domain operations, based on representing target elements and their dependencies as a network.

**Keywords:** targeting, all-domain operations, multi-domain operations, target selection, target-response matching, target engagement

---

## 1. Introduction

### 1.1 Motivation

A couple of years ago the Dutch tax authorities, several large banks, and other financial organisations suffered a series of heavy DDoS attacks over several days. This happened just after the Dutch intelligence agency AIVD had announced that it had penetrated the Russian Cozy Bear hacking group. The immediate thought was that the DDoS attacks were the Russian retaliation to this announcement, but investigation showed that an 18-year-old Dutch student was responsible (Volkskrant, 2018). The Dutch police were then faced with a choice: hack back or attempt to arrest him? They chose to send in an arrest team reportedly disguised as pizza couriers, enabling the seizure of the hacker's laptop as court evidence. In short, they responded kinetically to a cyber attack.

Combinations of cyber and kinetic action can also be found in military operations. In September 2007, the Israeli Defence Forces employed such a combination in Operation Orchard when they bombed a suspected nuclear reactor at Al Kibar, Syria (Fulghum, Wall & Butler, 2007). To hide the ongoing air raid, the Israelis reportedly took over the Syrian air defence system using network attack techniques, feeding the Syrian operators a false sky picture. In August 2008, the Russo-Georgian war began with online attackers assaulting Georgian websites to deny the Georgian government communications with its citizens and the outside world (Hollis, 2011). Cyberspace operations were synchronised with combat activity in the physical realm. In May 2019, the Israeli Defence Forces bombed and partially destroyed a building in Gaza, alleging that it was the base of an active Hamas hacking group (Newman, 2019).

The US Department of Defense (DoD) officially added the cyberspace domain to the existing four domains of land, sea, air, and space in 2011, with NATO following in 2016. Cyberspace is fundamentally different to the other four domains (Seebeck, 2019). The original four are natural physical domains, with material objects that can only be in one place at a time, movement being subject to inertia, and action and effects being kinetic. Cyberspace is a man-made domain, with virtual objects that may exist as multiple identical copies, inertia-less

movement, and cyber action and effects. There are close links between the virtual objects and the physical domains. Virtual objects can only exist in physical hardware, collectively known as infrastructure. Many virtual objects represent (attributes of) material objects, as when the identity of a computer is represented as an IP address.

Because of these differences, cyber operations doctrine (e.g. US Joint Publication (JP) 3-12 *Cyberspace Operations* (US DoD, 2018)) has been kept separate from doctrine for operations in the physical domains (e.g. US JP 3-0 *Joint Operations* (US DoD, 2017)). At present, cyber operations are often organised functionally (e.g. as in the US Cyber Command), rather than the geographically, as for the physical domains. The separation between cyber and physical domains is also found in the scientific literature on cyber warfare. Articles and papers invariably assume that operations will take place only in the cyberspace domain.

However, the emerging concept of all-domain operations (ADO) (a.k.a. multi-domain operations) (Underwood, 2020) (Economist, 2021) is domain-agnostic. Information from sensors in any domain may be used to select targets and to task resources from any domain to engage the selected targets. This implies that targeting must make choices between cyber and kinetic actions, confronting the differences between cyberspace and the physical domains. This paper focuses on this choice, which has been addressed neither in military doctrine nor in the scientific literature.

## **1.2 Purpose, scope, and paper structure**

The purpose of this paper is to propose an approach for choosing between cyber and/or kinetic action as part of the targeting process in all-domain operations. The focus is on target development and matching own capabilities, primarily in military operations. All-source intelligence is assumed to be available.

The paper is structured into six sections. After this introductory section, Section 2 reviews the relevant doctrine on targets, the targeting process, and targeting in cyberspace. Section 3 outlines the available literature on targeting in all-domain operations. Section 4 analyses a number of case studies, looking at the choices that the targetters are likely to have made between kinetic and cyber action. Section 5 proposes an approach to extending target system analysis to exploit dependency relationships in a network of target elements. Section 6 draws conclusions and makes recommendations for further work.

## **2. Related doctrine**

### **2.1 Targets**

Military doctrine defines a target as “an entity (person, place, or thing) considered for possible engagement or action to alter or neutralize the function it performs for the adversary” (US DoD, 2013, p.I-1). The physical, functional, cognitive, environmental, and temporal characteristics of each target form the basis for detection, location, identification, classification, analysis, engagement, and assessment. Targets can be categorised as shown in Table 1.

**Table 1:** Categories of military targets (US DoD, 2013), p.I-1 & -2)

<b>Category</b>	<b>Description</b>
Facility	A geographically-located physical structure, group of structures, or area, e.g. building
Individual(s)	A person or persons, e.g. combatants or suspects in law enforcement
Virtual	An entity in cyberspace, e.g. data, application, user or e-mail account
Equipment	A device, e.g. weapon, (fighting) vehicle, ICT hardware
Organisation	A group or unit, e.g. criminal gangs, extremist or terrorist cells, or military forces

As the definition of a target emphasises, these categories are entity-oriented. Relationships have a subordinate place in the categorisation of target characteristics. For example, dependencies on raw materials, personnel, energy, water and command & control (C2) are a sub-category of environmental characteristics (US DoD, 2013, p.I-2 to I-5).

According to Nisbett (2019), entity orientation is typical for the Western way of thought. Asian cultures (e.g. Chinese) think primarily in terms of relationships, with entities being subordinate. Nisbett’s insight is key to our proposed approach, in which entities are represented as nodes and relationships as links (a.k.a. arcs) in a network.

## 2.2 Targeting process

In military doctrine, the targeting process is an iterative cycle of six phases (US DoD, 2013, p.II-3):

- 1. *End state and commander's objectives*. In phase 1, the commander gains an understanding of the desired end state, objectives, effects, and tasks developed during operation planning.
- 2. *Target development and prioritization*. Phase 2 consists of three steps:
  - a) *Target system analysis*. Target systems are decomposed into components, which are then decomposed further into target elements, as shown in Figure 1.
  - b) *Entity-level target development*. The characteristics of each target element are determined.
  - c) *Target list management*. The list of targets is updated as they are vetted, validated, nominated, and prioritized.
- 3. *Capabilities analysis*. Phase 3 involves matching own capabilities against each target element to determine the options for engagement.
- 4. *Commander's decision and force assignment*. In phase 4, the commander decides which of the options will be selected, and assigns his/her own forces accordingly.
- 5. *Mission planning and force execution*. In phase 5, detailed plans for the selected options are generated, validated, and executed.
- 6. *Assessment*. Phase 6 is a continuous process that assesses the effectiveness of the operation. If the objectives have not been achieved, then the targeting cycle iterates.

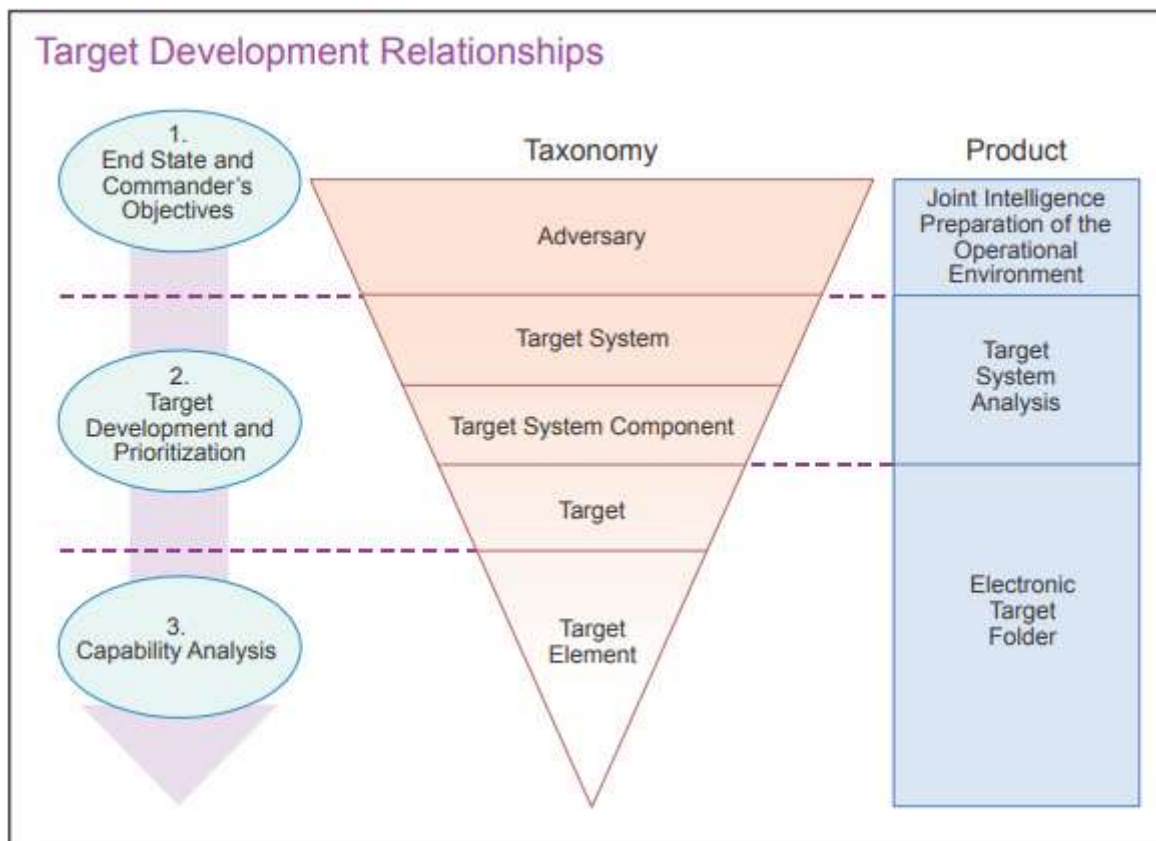


Figure 1: V-diagram: target system analysis (US DoD, 2013, figure II-3, p.II-6)

To illustrate target system analysis, the *Joint Targeting* doctrine manual, JP 3-60, depicts the results of analysing a typical (but fictitious) Air Defence Forces target system (US DoD, 2013, figure II-4, p.II-7). This system consists of two components: the Integrated Air Defense System (IADS) and airfields. The IADS's target elements are radar sites, anti-aircraft artillery, surface-to-air missile (SAM) sites, and Command, Control and Communications (C3). These elements include both physical and cyber entities, but lack a description of the relationships between them. For example, the C3 target element would be connected to all the other elements by communication

links. Moreover, the description omits to mention other elements outside this target system on which it is dependent, such as air traffic management (ATM), weather data, or electrical power. Attacking these external elements could render the IADS and airfield as inoperable as if they had been attacked directly.

### 2.3 Targeting in cyberspace

Cyberspace is the part of the information environment consisting of the interdependent network of ICT infrastructures and data (US DoD, 2018). It includes the Internet, communications networks, computer systems, embedded processors and controllers, and the information residing in and passing through the environment. Cyberspace operations (CO) are the employment of cyberspace capabilities aimed at achieving objectives in or through cyberspace, and may be defensive or offensive in nature.

In military doctrine, cyberspace is typically modelled as a set of distinct, yet interrelated layers. We adopt UK doctrine because this covers both physical and cyber domains. It recognises six layers (UK MoD, 2016): social, people, persona, information, network, and real, with the real layer having physical and geographical aspects. Table 2 shows how the target categories from Table 1 map into these layers.

**Table 2:** Mapping target categories into cyberspace layers

Layer (UK MoD, 2016)	Entities found in this layer	Target categories (US DoD, 2013)
Social	Organisations, groups & teams; human interaction; procedures; organisational culture	Organisation
People	Individuals (users, developers, administrators, maintainers)	Individual(s)
Persona	Accounts (user, e-mail, etc)	Virtual entities
Information	Data, applications & protocols; domain names; connections between nodes	
Network	Network nodes	Facility, Equipment
Real (with 2 aspects): Physical	Infrastructure; devices; cables; wireless & optical communication links	
Geographical	Locations of infrastructure, devices, cables, & communication links, and of individuals	

The relationships between entities often cross layers. For example, accounts (persona layer) are owned by individuals (people layer). Manned weapon systems, such as an armoured vehicle, ship, or aircraft, carry teams (social layer) of individuals (people layer), are constructed from physical devices (real layer, physical aspect), and can be found at a geographical location (real layer, geographical aspect). In the information layer, data may model devices in the real layer, e.g. data may represent an aircraft, with its identity, geographical location, altitude, speed, and direction of travel.

Doctrine on targeting in cyberspace is detailed in JP 3-12 *Cyberspace Operations* (US DoD, 2018), Chapter 4, Section 4 (pages IV-8 to IV-11). The main points are:

- The targeting process, target development, and validation are unchanged from JP 3-60.
- Three fundamental aspects of CO should be considered in targeting:
  1. Cyberspace capabilities are a viable option for engaging some targets.
  2. Cyber action may be preferable because it offers a low probability of detection (i.e. it is stealthy/covert) and/or results in no physical damage (i.e. the effects are temporary or reversible).
  3. Higher-order effects on cyber targets may provoke the adversary to retaliate.
- Cyberspace capabilities and targets have some unique characteristics, as follows:
  - *Targeting in and through cyberspace:*
  - CO effects may cross geographic boundaries.
  - CO may have unanticipated effects.

- CO must be closely coordinated with other units, with intelligence agencies, and with inter-agency and multi-national partners. This is because they may be planning or conducting their own operations within the same area of cyberspace.
- CO requires a Command & Control capability that can operate at the CO tempo and can rapidly integrate other stakeholders affected by these operations.
- The challenge in CO is to identify, correlate, coordinate, and deconflict activities across the following layers:
- *Real layer.* The real layer is the point of reference for determining geographical location, which in turn determines whether an intended action is within the (physical) area of operations and under which jurisdiction it falls.
- *Information layer.* The logical network within the information layer provides an alternative view of the target, referenced from its logical position in cyberspace. This is the first point at which the mapping to the physical domains may be lost. Targeting in the information layer requires the logical identity and logical access path to the target in order to have a direct effect.
- *Persona layer.* The persona layer contains the logical identities of individuals, groups, and organisations. These characteristics are needed for positive identification and attribution. Personas can be complex, with elements in many logical locations, but often not mapped to a single physical location. Links to entities in the real or information layers may be needed for engagement. Significant intelligence collection and analysis are required for effective targeting of a persona.
- *Access to targets in cyberspace:* Access to targets in cyberspace is developed through cyber intelligence collection. In some cases, remote access is not possible (e.g. Stuxnet), and close proximity may be required. All target access efforts in cyberspace require deconfliction with intelligence agencies and a consideration of potential intelligence gain/loss concerns. If direct access to a target is unavailable or undesired, indirect access via a related target may be feasible.
- *Target nomination:* Cyber-specific characteristics may be needed to understand how the target is relevant to the commander's objectives. Such characteristics are also needed to match an appropriate capability to engage the target.

Smart (2011) says that, despite the disparities between cyber and kinetic operations, cyber warriors are fundamentally the same as their kinetic counterparts. Both rely on their knowledge of the domain, operational environment, and weapon system capabilities, and can apply the same legal principles and military doctrine. He recommended the following modifications in targeting doctrine, none of which have yet been implemented in JP 3-12 (US DoD, 2018):

- Introduce the concepts of area of operations in cyberspace and an adversary's cyber centre of gravity.
- Provide an overview on how to conduct collateral damage estimation and battle damage assessment in cyberspace.
- Provide tactics, techniques, and procedures for identifying civilian and hostile websites and for tracing potential second- and higher-order effects and their likely geographical location.
- Recognising additional considerations for time-sensitive targeting in and through cyberspace.
- Distinguish between offensive and defensive cyber targeting, particularly as to the certainty of attribution needed.

### **3. Targeting in all-domain operations**

The current concept for operating over multiple domains is known as *joint operation*. Since almost all nations have one military service for each domain, units are drawn from each service to form a joint task force. The overarching military mission is decomposed by domain, which each unit planning operations in its own domain. While targeting may use intelligence from all domains ("all-source intelligence") as its input, one unit is rarely authorised to assign resources from another service. Instead, the set of plans must be coordinated to synchronise execution.

By contrast, in all-domain operations (ADO) (a.k.a. multi-domain operations), operational planning – and thus targeting – is domain-agnostic (Underwood, 2020) (Economist, 2021). The sensors providing intelligence on potential targets may be drawn from any domain. Likewise, the weapon systems that engage the targets

(“shooters”) may also come from any domain. All sensors and weapon systems are connected to a tactical network, enabling sensor-to-shooter peer-to-peer communication chains to be set up in seconds. ADO has a number of advantages:

- Vulnerable control centres, that might be eliminated by a single well-aimed missile, do not play a role in the sensor-to-shooter chains.
- A sensor or shooter that is failing can be readily swapped out and replaced by another.
- Strikes can be ordered in seconds, rather than minutes or hours as at present.

The ADO concept has reached the stage of experimentation. For example, in September 2020 off the coast of California, army artillery fed by instructions from air force sensors shot down a cruise missile in a response described as “blistering” (Economist, 2021). In the ABMS 3 experiment, held during Exercise Valiant Shield in the Pacific, also in September 2020, USAF KC-46 aircraft, US Navy vessels, and a ground station at Anderson Air Force Base, Guam, tested real-time communications outside the range of existing networks (Underwood, 2020).

Targeting in ADO follows the same targeting policy, principles and process as defined in JP 3-60 (US DoD, 2013). However, because ADO is domain-agnostic, there will be times when targetters must choose between assigning a cyber resource or a resource from one of the physical domains. In some cases, the best choice might be a combination of cyber and kinetic actions, as in the Dutch hacker, Hamas bombing, Operation Orchard, and Russo-Georgian cases. Targetters will then be confronted with the differences of targeting in cyberspace identified in JP 3-12 (US DoD, 2018) and by Smart (2011).

#### **4. Case studies**

Several cases have been analysed to see what choices the targetters will have made and why.

##### **4.1 Operation Orchard, 2007**

The choice facing the Israeli targetters was to destroy the Syrian air defence system kinetically or to deceive it using network attack techniques (Fulghum et al, 2007)<sup>1</sup>. A kinetic attack would have not only alerted the operators, but also the strong anti-aircraft defences around the nuclear reactor itself. By choosing cyber action, the operators would remain unaware of the impending air raid. This was also advantageous strategically (Makovsky, 2012). The Israeli targetters judged that the Syrian government would be unable to accuse the Israelis of the covert attack, without the shame of having to admit that North Koreans were building a nuclear reactor on Syrian territory, funded by Iran. Events proved this judgement to be correct.

##### **4.2 Russo-Georgian war, 2008**

The combination of cyber and kinetic action in the Russo-Georgian war is similar to Operation Orchard. The Russian targetters had a choice of cyber or kinetic action to degrade Georgian government communications (Hollis, 2011). If they had chosen kinetic action, this would have alerted the Georgians, resulting in an early kinetic response. By contrast, cyber action would delay the Georgian response until the Russians had already started seizing Georgian territory. Deniability was aided by employing militias, criminals, and individual hackers.

##### **4.3 Hamas hacker bombing, 2019**

When faced by the Hamas hacker group, the Israeli Defence Forces had a choice of hacking back or responding kinetically (Newman, 2019), like the Dutch police in 2018. Hacking back would have only disrupted the ICT equipment, allowing the hackers to resume their activities with new equipment. Unlike the Dutch police, the Israelis did not have to present any evidence in a court of law, and so they chose to respond destructively.

The Israelis then had a second choice: whether to send in special forces to capture the Hamas hackers, or to bomb the building together with its contents. The latter option stopped the hackers’ activities permanently, without the risk involved in a capture mission. Moreover, the bombing sent a clear message to other Hamas fighters.

---

<sup>1</sup> For clarity, we simplify what actually happened: network attack delivered by specialised aircraft.



#### 4.4 Stuxnet, 2008

Although Stuxnet was purely a cyber weapon (with second-order kinetic effects), it is instructive to compare the targetters’ choices. As in Operation Orchard and the Russo-Georgian war, Stuxnet had two “warheads” (Zetter, 2014): one aimed at kinetic self-destruction of the centrifuges, and the other at deceiving the operators by providing them with a false picture of normality. The targetters again had the choice of kinetic or cyber action. To destroy the underground centrifuges kinetically, the targetters would have had to use either a massive penetrating bomb or a nuclear weapon. This could not be done without telling the world who the attackers were. The use of a nuclear weapon would lead to the attackers being globally condemned. Instead, they chose cyber action for stealth and deniability.

#### 4.5 Ukrainian power grid, 2015

In December 2015, the Ukrainian electrical power grid was hit by a CO consisting of five different attacks (Lee, Assante & Conway, 2016). The main attack targeted the power breakers and the power distribution to black out large areas (second-order physical effects). One second attack erased the backup system, and a third prolonged the black-out by overloading the telephone numbers for reporting electrical faults. The fourth attack hit the operators’ workstations, deleting the operating system and making them permanently unusable. The fifth targeted the substations by overwriting the serial-to-Ethernet code so that they could not be operated remotely, delaying recovery and repair.

Conducting the operation kinetically would have required Russian military forces to enter Ukraine in strength, leading to global condemnation. Instead, all of these attacks were conducted in or through cyber space, with three attacks having second-order kinetic effects.

#### 4.6 Summary

The results of this analysis are summarised in Table 3. They indicate that cyber action is suited to deception, delay, and access denial effects on organisation and virtual entities in the social, persona, and information layers. By contrast, kinetic action is suited to seizure, capture, and destruction of physical entities in the people, network, and real layers. Where kinetic action must be ruled out (e.g. for legal or strategic reasons), then cyber action should be sought that has a second- (or higher-) order kinetic effect.

**Table 3:** Results of analysis of case studies

Case	Trigger	Response chosen	Rationale
Operation Orchard	physical (WMD threat)	cyber (deceive operators) kinetic (destroy reactor)	Delay alerting air defence & deniability (operational & strategic)
Russo-Georgian war	physical (Georgian action)	cyber (degrade communications) kinetic (seize territory)	Delay Georgian response & deniability
Hamas hackers	cyber attack	kinetic (destroy building & equipment, and kill hackers)	Stop hacking permanently & send clear message
Stuxnet	physical (WMD threat)	cyber (deceive operators) kinetic (destroy centrifuges)	Kinetic unacceptable. Cyber stealthy & deniable
Ukrainian power grid	physical (Russian aggression)	Cyber operation (5 attacks: shut off electrical power to large areas, deny access to fault reporting, delay recovery & repair)	Kinetic unacceptable. Cyber deniable (but overt)

### 5. Proposed approach

#### 5.1 Network representation

Our proposed approach is based on representing targets as a network of entities (nodes) and inter-entity relationships (links). This network representation is known in the intelligence community as a target network model (TNM) (Clark & Mitchell, 2016). After intelligence collection and analysis, the target network generally comprises entities in the social, people, and real layers. For example, the network may describe the structure of a formal or informal organisation (social layer), individuals with their family and social relationships (people layer), or physical devices with their geographic locations (real layer). If cyber intelligence is available, there may

also be information on virtual entities (information and persona layers) and on ICT hardware (network and real layers).

### 5.2 Dependency network analysis

Our proposed approach focuses specifically on the dependency relationships in the network to identify chains of dependency relationships. For example, one entity may be dependent on electrical power generated by another (e.g. a power supply), or one entity may be dependent on information collected by another (e.g. a sensor). This is known as *dependency network analysis* (Drabble, 2014).

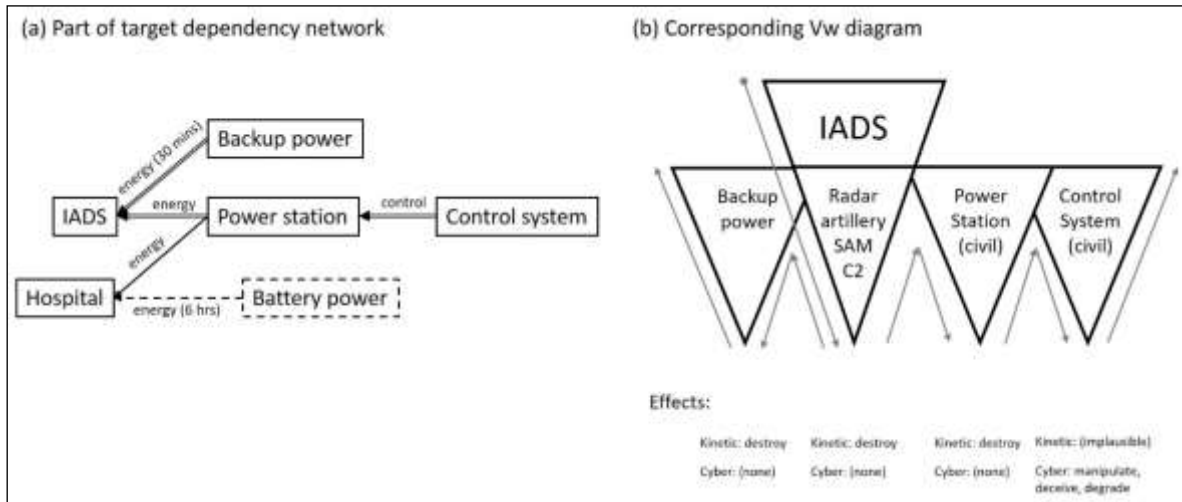


Figure 2: Example target dependency network (a) and corresponding Vw diagram (b)

### 5.3 Illustrative dependency network

To illustrate our proposed approach, we expand on JP 3-60’s example of an IADS (US DoD, 2013, figure II-4, p.II-7). To demonstrate dependency relationships and the cyber-versus-kinetic choice we add some external (civilian) entities: a power station and its computer-based control system, and a hospital which may or may not have emergency battery power. We assume that intelligence-gathering on the IADS has shown it gets its energy from the power station, but also has backup power for 30 minutes. The hospital also gets energy from the power station. The relevant part of the dependency network is shown in Figure 2(a).

### 5.4 Description of analysis

The first step in the analysis is to select potential targets from the TNM that, when engaged, would achieve the commander’s objectives. For each potential target, an initial V diagram is constructed and an electronic target folder is opened containing all the target’s known characteristics. In the illustrative case, we select the IADS component, resulting in the large V in Figure 2(b). The downward arrow alongside the large V symbolises decomposition of the IADS component into its target elements (phase 2: target development). Upward arrows symbolise matching own capabilities to the target (phase 3: capabilities analysis), with the choice between kinetic and cyber action being made as outlined in Section 4.6.

The second step exploits the dependency relationships in the network. Observing that the IADS is a heavily-defended, dispersed target, the targetter looks for other entities on which the IADS is dependent, i.e. upstream in the dependency network. The power station is one possibility, because cutting off power to the IADS would make it easier to attack. A small v is added to the right of the large V, symbolising the power station. A second possibility is the backup power, because if the IADS could not obtain energy from the power station then it would be dependent on backup power, although this would be limited to 30 minutes. We add another small v to the left of the large V.

However, the power station is civilian, meaning that it cannot lawfully be targeted. The analysis iterates to look further upstream in the dependency network. The targetter observes that the power station is dependent on its control system, consisting of hardware and software. While a kinetic attack on the hardware would be unlawful,

a cyber attack on the software might be acceptable if it was temporary, reversible, and deceived the control system's operators. Another small v is added to the right, symbolising the control system.

The analysis iterates again, this time downstream, to check if there are other entities also dependent on the power station. The network shows that there is a hospital that depends on energy from the power station. Since international law forbids attacks on hospitals, its dependence would seem to rule out shutting down the power station, even if the attack was temporary and reversible. However, the targetter knows that hospitals often have their own emergency power supply. Since this is not definitely shown in the dependency network (dashed lines), the targetter tasks the intelligence agency to confirm whether or not the hospital has emergency power. After some delay, intelligence gathering shows that the hospital has emergency battery power good for six hours. This knowledge permits a cyber attack on the power station's control system.

Finally, the targetter concludes from the V diagram that a combination of kinetic and cyber attack is needed to bring down the IADS. The control system must be attacked via cyberspace to shut down the power station and deceive its operators. The IADS's backup power must be destroyed by kinetic means to make the whole IADS inoperative. The IADS can then be attacked kinetically to destroy it, after which the power station can be switched on again. This operation must be completed in under six hours to avoid exhausting the hospital's emergency battery power.

### **5.5 Automating the proposed approach**

While our proposed approach enables targetters to consider both kinetic and cyber options, as well as their combination, it does have the disadvantage that the targeting process iterates, thus taking longer. This disadvantage can be mitigated by automating the approach using AI techniques, as demonstrated by Drabble (2014). AI-based dependency analysis has been implemented and demonstrated to DARPA for at least two scenarios: an IADS and an insurgent cell.

## **6. Conclusions and further work**

Targeting is the process of selecting and prioritizing targets and matching the appropriate response to them. The process is the same whether the response is cyber, kinetic, or some combination. However, there are major differences between cyberspace and the physical domains (i.e. land, sea, air, and space). The emerging concept of all-domain operations requires targetters to weigh actions and effects across all five domains. Instead of engaging physical targets only by kinetic actions and cyber targets only by cyber means, they must consider both cyber and kinetic action against any target, including second- and higher-order effects. Existing military doctrine and the scientific literature give no guidance on this.

This paper attempts to fill the gap by proposing an approach for choosing cyber and/or kinetic action as a part of the targeting process. The approach is based on regarding targets as a network of entities with dependency relationships between them. Where a target cannot be engaged, then another target, perhaps in another domain, on which the original target is dependent may be instead. Stuxnet is the classic example where physical entities (namely centrifuges) could be destroyed as a second-order effect of a cyber attack. A disadvantage of the proposed approach is that dependency network analysis is more time-consuming than the existing process, but this can be mitigated by automating it using AI techniques. AI-based dependency analysis has been demonstrated.

The main contribution of this paper is to show how the choice might be made between cyber and/or kinetic action as a part of the targeting process. The key limitation is that, while AI techniques have been applied to dependency network analysis, this has been in generating plans, rather than targeting.

Further work is needed to demonstrate that AI-based dependency analysis can be applied to targeting. Moreover, targeting doctrine needs to be modified as Smart (2011) suggests.

## **References**

Clark, R.M. & Mitchell, W.L. (2016). *Target-Centric Network Modeling: Case studies in analyzing complex intelligence issues*. SAGE Publications Inc., Thousand Oaks, CA.

**Tim Grant and Harry Kantola**

- Drabble, B. (2014). Modeling C2 Networks as Dependencies: Understanding What the Real Issues Are. In Grant, T.J., Janssen, R. and Monsuur, H. (eds.) *Network Topology in Command and Control: Organization, Operation, and Evolution*. IGI Global, Hershey, PA, 125–151.
- Economist. (2021). America's approach to command and control goes peer to peer. *Science & technology, Economist*, 9 January 2021.
- Fulghum, D.A., Wall, R. & Butler, A. (2007). Cyber-Combat's First Shot: Attack on Syria shows Israel is master of the high-tech battle. *Aviation Week & Space Technology*, 26 November 2007, 28-31.
- Hirsch, C. (2018). Collateral damage outcomes are prominent in cyber warfare despite targeting. In Chen, J.Q. and Hurley, J.S. (eds.), *Proceedings, 13<sup>th</sup> International Conference on Cyber Warfare and Security (ICCCWS 2018)*, 281-6.
- Hollis, D. (2011). Cyberwar Case Study: Georgia 2008. *Small Wars Journal*, Small Wars Foundation, 6 January 2011.
- Lee, R., Assante, M. & Conway, T. (2016). Analysis of the cyber attack on the Ukrainian Power Grid. E-ISAC, Washington DC.
- Makovsky, D. (2012). The silent strike. *Annals of war, New Yorker*, 10 September 2012.
- Moore, D. (2018). Targeting Technology: Mapping military offensive network operations. *Proceedings, 10<sup>th</sup> international conference on Cyber Conflict (CyCon 2018)*.
- Newman, L.H. (2019). What Israel's strike on Hamas hackers means for cyberwar. *Wired*, 6 May 2019.
- Nisbett, R.E. (2019). *The Geography of Thought: How Asians and Westerners think differently*. Nicholas Brealey Publishing, London & Boston.
- Seebeck, L. (2019). Why the Fifth Domain is Different. *The Strategist*, Australian Strategic Policy Institute, 5 September 2019.
- Smart, S.J. (2011). Joint Targeting in Cyberspace. *Air & Space Power Journal*, Winter 2011, 65-75.
- Underwood, K. (2020). Holding the line for joint all-domain command and control. *The Cyber Edge*, 1 November 2020.
- UK MoD. (2016). *Cyber Primer*. 2<sup>nd</sup> edition, Development, Concepts & Doctrine Centre, MoD Shrivenham, UK, July 2016.
- US DoD. (2013). *Joint Targeting*. US DoD Joint Publication 3-60, 31 January 2013.
- US DoD. (2017). *Joint Operations*. US DoD Joint Publication 3-0, 17 January 2017.
- US DoD. (2018). *Cyberspace Operations*. US DoD Joint Publication 3-12, 8 June 2018.
- Volkskrant. (2018). 18-jarige jongen opgepakt in verband met DDoS-aanvallen op Belastingdienst. *Volkskrant*, 5 February 2018. [In Dutch: 18-year-old youth arrested in connection with DDoS attacks on the Tax Service.]
- Zetter, K. (2014). *Countdown to Zero Day: Stuxnet and the launch of the world's first digital weapon*. Crown Publishers, New York.
- Zetter, K. (2016). Inside the cunning, unprecedented hack of Ukraine's power grid. *Wired*, March.

# Computer Aided Diagnostics of Digital Evidence Tampering (CADET)

Babak Habibnia, Pavel Gladyshev and Marco Simioni

DFIRE, University College Dublin, Ireland

[Babak.Habibnia@ucd.ie](mailto:Babak.Habibnia@ucd.ie)

[Pavel.gladyshev@ucd.ie](mailto:Pavel.gladyshev@ucd.ie)

[marco.simioni@ucdconnect.ie](mailto:marco.simioni@ucdconnect.ie)

DOI: 10.34190/EWS.21.059

**Abstract:** The tampering of the digital crime scene has become more common. When tampering behaviour is successful, it does not leave a trace of either the incriminating evidence or the act of tampering and the digital evidence that digital investigators seek will be absent. The research into the automatic detection of digital evidence tampering has been ongoing for over 13 years. Many approaches had been proposed, but the practical tools for automatic or semi-automated detection of evidence tampering are still missing. Due to the complexity of real-world computers and the differences between software installed on different computers automatic analysis is hard. A similar problem exists in medical imaging. Despite the common grand design, every human is unique and complex, and it is hard to come up with the exact rules for detecting lesions in medical images. Visualization for forensic analysis of the data stored on a specific device has received much less attention, while the use of visualization for detection of digital evidence tampering is virtually unexplored. This paper proposes, for the first time, a semi-automated approach based on visualization of relevant data properties, helping human investigators to detect digital evidence tampering and anomaly. This is analogous to computer-aided processing of medical X-Ray images that enhance the visibility of lesions facilitating easier detection by a doctor. This paper aims to identify data tampered features on the digital devices, then find suitable visualization to display identified data tampered features for investigators. One of the outstanding features of the approach proposed in this paper for detecting digital evidence tampering is its malleability. It can easily apply to any specific or whole part of data in the digital devices, visualize, and reveal offender concealment behaviour concerning the detection of evidence tampering.

**Keywords:** cybercrime, cybersecurity, digital evidence tampering, digital forensics, anti-forensics, visualization

---

## 1. Introduction

Any action by a user on a computer, whether it's surfing the internet, communication or file storage, affects data stored in the computer. Analysis of computer data can often help to determine when, where, and how a crime has been committed. Digital forensics is a branch of forensic science dealing with inspection, extraction, and analysis of computer data as evidence in litigation. Currently, when the digital investigator is faced with any type of digital evidence tampering behaviour, they must look for present forensic tools or methods in an ad hoc manner, for instance, ask peers, search tool sources to look for similarity evidence tampering, review publications. Even if a given problem has been found, there is currently a delay in disseminating a new tool or publishing a new method. So far, most computer forensics tools offer capabilities such as imaging, analysis, viewing, and reporting. They are unable to present a visual overview of all data found on a piece of media especially when evidence tampering occurred. This paper aims to detect digital evidence tampering and it has been divided into two main parts. The first part deals with identifying data tampered features on the digital devices which focusing on the six key tasks, identifying anti-forensics tool features, used dataset, evidence tampering action, comparison, result, and exploring automated method. The second part deals with finding suitable visualization to visualize identified data tampered features for investigators, focusing on the two key tasks, designing visualization-parallel Coordinates, and result. For instance, visualize PC1, PC2, PC3, PC4, and PC1 (tampered) behaviour in a normal way of usage, base on the identified data tampered features.

## 2. Methodology

This section covers the unique explored methodology for detecting digital evidence tampering. It is divided into the two following main sub-sections which present, the methodology was taken and the result.

### 2.1 Identifying data tampered features (file + location)

In order to identify data tampered features, it will be necessary to carry out the following steps.

2.1.1 Identifying Anti-Forensics (AF) tool features

The purpose of this part is to identify anti-forensics tool features and describe them where and what the impact on the data after is using. Any action by a user on a computer (or digital devices), affects data stored in the computer. Different AF tools have a varying degree of effectiveness. In a positive aspect, AF tools are used to sanitize user activity and conversely can lead to intentional tampering (deliberate manipulation). By reviewing three common and free available AF tools, Windows Eraser (GOG, 2002), Track Eraser (Acesoft, 2001) and CCleaner (Dimmick, 2005) it stands out each of the AF tools has a unique GUI (graphic user interface) and particular features. Table 1 shows a comprehensive overview of the AF tool’s features based on the Windows Operating System.

Table 1: Anti-Forensics tool’s features

Features & AF Tools		Features																														
		Erase History Data	Cache	Cookies	History	Auto-Complete Memory	Typed URL	index.dat	Win Temp Folder	Run & Search History	Open/Save History	Recent Documents	Removes Unused Files	Clean Traces Online Activities	Memory Dump	Internet History	Error Reporting	Jump List	Registry Cleaner	Recycle Bin	Clipboard	Log Files	DNS Cache	Super Cookie	Unwanted Traces	Password	AC/Key	Adobe	MSN	Yahoo Messenger	Microsoft Office	Erase online & Offline (only IE)
AF Tools	Tracks																															
	Erase Pro	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	Windows Eraser	-	-	✓	-	✓	-	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	
	CCleaner	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

2.1.2 Used dataset

In this study, the authors used 4 virtual machine disk images from the DFIRE lab (DFire, 2013) to simulate an original intellectual property theft that occurred in a company in 2015 for the experiment. The disk images included the following specification:

- Windows 7.
- Over 93,012 (test case before tampering purpose) and 94,111 (test case after tampering) records in 35 columns.
- Installation of two additional commonly used browsers: Chrome and Firefox.
- Using all three browsers Internet Explorer, Chrome and Firefox.
- Running several programs.

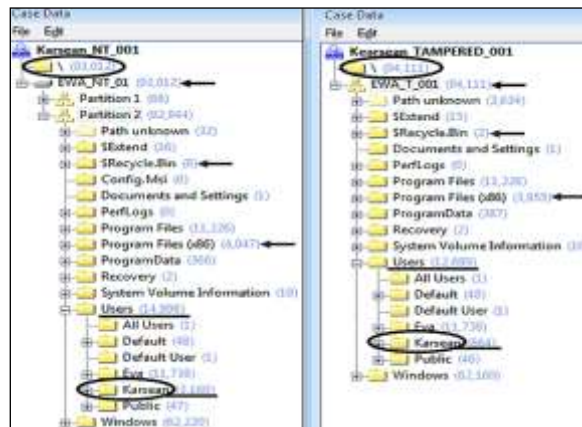
2.1.3 Evidence tampering action

This step was carried out using the CCleaner software (Dimmick, 2005) which is available in both free and paid format, to create a tampered image for comparison (next step) in the designed test environment. According to the obtained result shown in Table 1 Anti-Forensics Tools Features, the CCleaner cover all features and more than it was expected for testing in the sample case (dataset). The standard dataset base of the Win 7 which provided by the DFIRE lab (DFire, 2013) was set up for tampering on the test environment. The CCleaner was prepared for tampering purpose, conducted till complete its deletion and wiping. Then the image was taken using FTK imager tool from tampered data and examined with X-Ways (“X-Ways Forensics,” 2002) suite of computer forensics tool. At first, by looking and analyzing tampered data using the X-Way, there was no knowledge of how and where exactly CCleaner affected the data. After that, considering the AF tool’s features (Table 1) and review data in-depth (folders, subfolders, and files), didn’t help to distinguish the effects of tampering (or manipulating). It only shows the number between the brackets in front of each folder.

2.1.4 Comparison

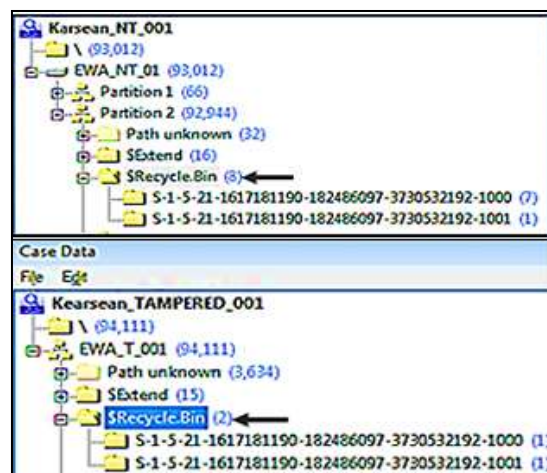
This step follows from the previous step 2.1.3, which examined the comparison between cases. This time, the image was taken using the FTK imager (“AccessData,” 2010) tool from the Karsean case without conducting a CCleaner tool for tampering purpose (non-tampered). Using a similar method and a tool for analyzing data showing exciting results in the numbers are included as bracketed points. That encouraged to carry out a

comparison and see a difference between Tampered and Non-Tampered Data. At first glance, a comparison of two results reveals a big difference between the number of the files in Tampered and Non-Tampered data as indicated in Figure 1.



**Figure 1:** Comparison between a number of the files in a tampered and non-tampered dataset (Kearsean\_Tampered and Kearsean\_NT cases)

According to the obtained results shown in Table 1 (Description of Anti-Forensics Features), in this step, we were looked for any individual folder, sub-folder, and file. For instance, in Figure 2, *Recycle.bin* included deleted file (Wikipedia, 2016) shows clearly the difference between the number of files between both systems before and after tampering (existing & absence). Another very interesting item could be the username folder, and subfolders, which contains very important files such as web browser application, cache, history, cookies, etc. concerning tampering purpose.



**Figure 2:** Comparison between a number of the files in a non-tampered and tampered dataset in Recycle Bin, 8 (NT) --> 2 (T) (Kearsean\_NT and Kearsean\_Tampered cases)

### 2.1.5 Result (identifying data tampered features)

As explained in the previous steps, the X-Way forensics tool was used to compare the difference between the number of the files in the specific folders (e.g. RecycleBin, Users etc.). In this Section, for further analysis and to identify data tampered features with a comparison between both systems, tampered and non-tampered data were separately exported in two individual CSV files. The non-tampered file contained 93,012 records (row) and 35 columns entries and the tampered file contained 94,111 records (row) and 35 columns. Figure 3 shows exported data which included the title of each column and the content of each row where files and location are allocated.

It was found the difference between non-tampered and tampered data by looking and comparing into each row and column. For instance, the results obtained from the comparison of the *Recycle.bin* in two exported files, as



shown in Figure 4 and Figure 5 indicates the difference between the files and also identifying the filename and the location of the tampered file into the dataset before and after tampering action.

Record update	Deletion	Int. cre/Attr.	Owner	Links	File count	1st sector	ID	Int. ID	Int. parent	Dimens.	SCN	Hash	Hash Set	Hash Categ.	Report	Comme Metadata
01/03/2009 12:38		AX	Karssen 5-5-21-2017181199-182486097-3730532192-1000		6408870	50807	70329	113548								
01/03/2009 12:38		AX	Karssen 5-5-21-2017181199-182486097-3730532192-1001		6417752	83136	83678	113548								

Figure 3: CSV generated and exported

\\\$Recycle.Bin\S-1-5-21-1617181199-182486097-3730532192-1000
\\\$Recycle.Bin\S-1-5-21-1617181199-182486097-3730532192-1000
\\\$Recycle.Bin\S-1-5-21-1617181199-182486097-3730532192-1000
\\\$Recycle.Bin\S-1-5-21-1617181199-182486097-3730532192-1000
\\\$Recycle.Bin\S-1-5-21-1617181199-182486097-3730532192-1000
\\\$Recycle.Bin\S-1-5-21-1617181199-182486097-3730532192-1000\SR9SQ1ZM.pptx
\\\$Recycle.Bin\S-1-5-21-1617181199-182486097-3730532192-1000\SRD8EBAL.docx
\\\$Recycle.Bin\S-1-5-21-1617181199-182486097-3730532192-1001

Figure 4: Recycle Bin non-tampered dataset (filename + location) in CSV

\\\$Recycle.Bin\S-1-5-21-1617181199-182486097-3730532192-1000
\\\$Recycle.Bin\S-1-5-21-1617181199-182486097-3730532192-1001

Figure 5: Recycle Bin tampered dataset (filename + location) in CSV

Finally, after examining and comparing all entries of data, we achieved very valuable and strong results are to discover/or identify data tampered features. Table 2, presents the remarkable result of identifying data tampered features (file + location) which solved the first problem as mentioned in this research paper earlier. This finding never identified before the digital forensics field. This finding was never identified before in the digital forensics field.

Table 2: Identifying data tampered features for detecting digital evidence tampering which was never explored before

Identifying Data Tampered Features (file + location)		
Data Tampered	Location of Data Tampered in Microsoft Window XP/7/8/10	
Recycle.bin	Windows XP (C:\RECYCLER" 2000/NT/XP/2003) Win7/8/10 (C:\Recycle.bin)	
History	Internet Explorer	IE6-7 (%USERPROFILE%\Local Settings\History\History.IE5) IE8-9 (%USERPROFILE%\AppData\Local\Microsoft\Windows\History\History.IE5) IE10-11(%UERPROFILE%\AppData\Local\Microsoft\Windows\WebCache\WebCacheV*.dat)
	Firefox	XP (%USERPROFILE%\Application Data\Roaming\Mozilla\Firefox\Profiles\ <random text&gt;.default\places.sqlite)<br=""></random> Win7/8/10 (%USERPROFILE%\AppData\Roaming\Mozilla\Firefox\Profiles\ <random td="" text&gt;.default\places.sqlite)<=""> </random>
	Chrome	XP (%USERPROFILE%\Local Settings\Application Data\Google\Chrome\User Data\Default\History) Win7/8/10 (%USERPROFILE%\AppData\Local\Google\Chrome\UserData\Default\)
Cache	Internet Explorer	IE8-9 (%USERPROFILE%\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5) IE10 (%USERPROFILE%\AppData\Local\Microsoft\Windows\Temporary Internet Files\ Content.IE5) IE11 %USERPROFILE%\AppData\Local\Microsoft\Windows\INetCache\IE)
	Firefox	XP (%USERPROFILE%\Local Settings\ApplicationData\Mozilla\Firefox\Profiles\ <randomtext&gt;.default\cache) </randomtext&gt;.default\cache)  Win7/8/10 (%USERPROFILE%\AppData\Local\Mozilla\Firefox\Profiles\ <randomtext&gt;.default\cache)< td=""> </randomtext&gt;.default\cache)<>
	Chrome	XP (%USERPROFILE%\Local Settings\Application Data\Google\Chrome\User Data\Default\Cache - data_# and f_#####) Win7/8/10 (%USERPROFILE%\AppData\Local\Google\Chrome\UserData\Default\ Cache\ - data_# and f_#####)
CO	Internet	IE8-9 %USERPROFILE%\AppData\Roaming\Microsoft\Windows\Cookies



Identifying Data Tampered Features (file + location)	
Data Tampered	Location of Data Tampered in Microsoft Window XP/7/8/10
Data Tampered	Explorer IE10 %USERPROFILE%\AppData\Roaming\Microsoft\Windows\Cookies IE11 %USERPROFILE%\AppData\Local\Microsoft\Windows\INetCookies
	Firefox XP %USERPROFILE%\Application Data\Mozilla\Firefox\Profiles\<random text>.default\cookies.sqlite Win7/8/10 %USERPROFILE%\AppData\Roaming\Mozilla\Firefox\Profiles\<randomtext>.default\cookies.sqlite
	Chrome XP %USERPROFILE%\Local Settings\Application Data\Google\Chrome\User Data\Default\Local Storage\ Win7/8/10 %USERPROFILE%\AppData\Local\Google\Chrome\UserData\Default\ Local Storage\
Session Restore	Internet Explorer Win7/8/10 %USERPROFILE%\AppData\Local\Microsoft\Internet Explorer\Recovery
	Firefox Win7/8/10 %USERPROFILE%\AppData\Roaming\Mozilla\Firefox\Profiles\<randomtext>.default\sessionstore.js
	Chrome Win7/8/10 %USERPROFILE%\AppData\Local\Google\Chrome\User Data\Default\ Files = Current Session & Current Tabs & Last Session & Last Tabs
Prefetch	XP (C:\%USERPROFILE%\Recent) Win7/8/10 ( C:\%USERPROFILE%\AppData\Roaming\Microsoft\Windows\Recent\ & C:\%USERPROFILE%\AppData\Roaming\Microsoft\Windows\Recent\AutomaticDestinatins & C:\%USERPROFILE%\AppData\Roaming\Microsoft\Windows\Recent\CustomDestinations)
Temporary	%USERPROFILE%\AppData\Local\Temp
Shortcut (LNK)	XP (C:\%USERPROFILE%\Recent) Win7/8/10 ( C:\%USERPROFILE%\AppData\Roaming\Microsoft\Windows\Recent\ & C:\%USERPROFILE%\AppData\Roaming\Microsoft\Windows\Recent\AutomaticDestinatins & C:\%USERPROFILE%\AppData\Roaming\Microsoft\Windows\Recent\CustomDestinations)
Recent Files	XP (C:\%USERPROFILE%\Recent) Win7/8/10 ( C:\%USERPROFILE%\AppData\Roaming\Microsoft\Windows\Recent\ & C:\%USERPROFILE%\AppData\Roaming\Microsoft\Windows\Recent\AutomaticDestinatins & C:\%USERPROFILE%\AppData\Roaming\Microsoft\Windows\Recent\CustomDestinations)
Jump List	Win7/8/10 ( C:\%USERPROFILE%\AppData\Roaming\ Microsoft\Windows\Recent\AutomaticDestinations\ ID numbers.automaticDestinations-ms) Win7/8/10 ( C:\%USERPROFILE%\AppData\Roaming\Microsoft\Windows\Recent\CustomDestinations\ ID numbers.customDestinations-ms)
RDP Usage	XP (%SYSTEM ROOT%\System32\config\SecEvent.evt) Win7/8/10 (c:\Windows\System32\winevt\logs\Security.evtx)
Event Log	XP (C:\Windows\System32\winevt\Logs) Win7/8/10 (C:\Windows\System32\winevt\Logs)

2.1.6 Exploring automated method for detecting digital evidence tampering, using Prolog Styla.

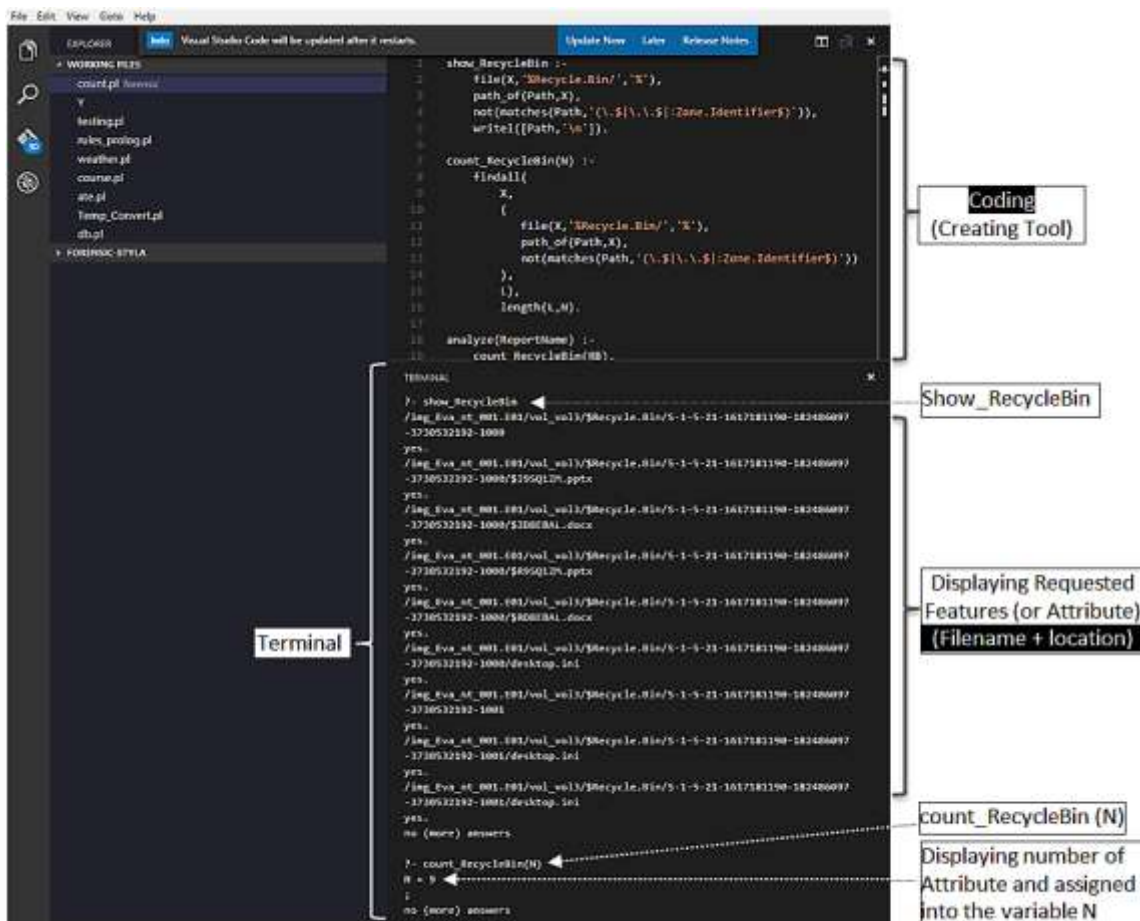
Following on the previous step 2.1.5, which outlined the significant result of the identifying data tampered features (file + location). The purpose of this step to create a reliable and flexible automated tool to display, count, and save the result of data tampered from a provided disk image to detecting digital evidence tampering, then export the result for Visualization. According to Table 2, it is almost difficult and impossible to identify data tampered features (file + location) manually to detect data tampering; especially when investigators with no prior knowledge/background of data. For this purpose, use Forensics Styla environment-Prolog in Scala with forensic extensions (Dfire, 2016). The styla is a lightweight implementation of a Prolog (Wikipedia, 1972) in Scala developed by Paul Tarau (Tarau, 2014). To examine data with Forensic Styla, it will be necessary to either create a new case or open an existing case file (for instance: *Karsean* as mentioned above). The following predicates are defined to do that: new case, open case, close case, deli case. Besides, it gives accessing/exploring the files into the image (test case) through the command line (terminal) when using a specific command. The new case was created in Forensics Styla and added the image (non-tampered or tampered) file was taken. Then, created a tool (or coding) to display, count, and save the resulting base on the investigator interest (partial or the entire

dataset), and exported it as a file for visualization, see *Table 3*. For instance, the authors were interested in some parts of features as a sample in this study. For instance, RecycleBin (RB), Internet History (IH), Cache (CSH), Cookies (COO), Restore Point (RP) and Prefetch (PR) features from *Table 2* for the experiment (count number of the files with considering their location) as exporting as shown in *Figure 6* for the visualization.

**Table 3:** Exported dataset among value in CSV format

	A	B	C	D	E	F	G
EVIDENCE	RB	HI	CSH	COO	RP	PR	
PC1	8	123	500	432	55	122	
PC1(Tamperd)	2	21	0	16	11	25	
PC2	25	175	515	531	64	135	
PC3	33	185	650	625	55	114	
PC4	46	191	569	571	35	163	

The significant reason for choosing Forensics Styla is that having both forensics tool and writing logic code interface on one front page instead of using two separate tools. It also uses a terminal for typing any query for further analysis. So far, this research paper focussed on identifying data tampered features (file+location) for detecting digital evidence tampering, which has achieved remarkable results and never been done in the computer forensics field *Table 1*. The following part which is the last part of this research study, will discuss visualization, The question that arises is, what is the best visualization to present obtained data in such an understanding and simple way for investigators?



**Figure 6:** Forensics Styla environment

## 2.2 Finding suitable visualization

Visualization allows for displaying an overview of all the data found on the piece of media. Teerlink and Erbacher (Teerlink and Erbacher, 2006) wrote in one sentence, what summarizes the problem:

'A great deal of time is wasted by analysts trying to interpret massive amounts of data that isn't correlated or meaningful without high levels of patience and tolerance for error. Visualization techniques can greatly aid forensic specialists to direct their search to suspicious file'.

It's easy to throw the data upon a bar chart or scatter plot in Excel, PowerPoint, slip it into a report, and convinced that it does the explaining. But that's a terrible shortcut. When the report is over, and the only thing left behind is the report and no-one will have a knowledge what the chart was trying to communicate or say. There are so many chart types, styles, and methods of presenting data that can be confusing and hard to pick the right chart type for obtained data in this paper. The following step discusses designing a suitable visualization.

2.2.1 Designing visualization – parallel coordinates

As it was pointed out in the introduction, the aim of this paper divided into two main parts: Identifying data tampered features and finding suitable visualization for detecting digital evidence tampering. Having identified the data tampered features in the previous part with details, and the maximum size of data is 21 as described in Table 2. Now, the authors will discuss finding suitable visualization. It was given a set of data points  $D = \{pc_i\}$  where every point  $pc_i$  has an n-dimensional vector of attributes  $(a_1^i, \dots, a_n^i) \in A^n$  defined on some domain A (e.g., See the file name in Table 2). Such a dataset called *multivariate* with several attributes or variables per data point as shown in Table 4. In this paper, the authors are interested in examining the distribution, correlation, and comparison of the individual values in the various dimensions and giving the overall distance between the data points. One technique that allows authors to perform such visualization is the *Parallel Coordinates* (Bostock, 2020; Chang, 2013; Davies, 2016; Inselberg, 1997; Tableau, 2015).

Table 4: Describing dataset in multivariate visualization

$D = \{pc_i\}$ then $D = \{pc_1, pc_2, pc_3, pc_4, pc_5\}$	<b>Data Points</b> (e.g., PC's, laptop and etc.)
$A = \{RB, IH, CSH, COO, RP, PR\}$ 1 2 3 4 5 6 (number of attributes)	<b>Attributes</b> (identified data tampered)
$(a_1^i, \dots, a_n^i) \in A^n$ then $(a_1^i, \dots, a_6^i) \in A^6$ and $(a_{RB}^i, a_{IH}^i, a_{CSH}^i, a_{COO}^i, a_{RP}^i, a_{PR}^i) \in A^6$ for $pc_i$	
$(a_1^i, \dots, a_n^i) \in A^n$ then $(a_1^i, \dots, a_6^i) \in A^6$ and $(a_{RB}^i, a_{IH}^i, a_{CSH}^i, a_{COO}^i, a_{RP}^i, a_{PR}^i) \in A^6$ for $pc_i$	

Parallel coordinates are one of the most common ways of visualizing and analyzing multivariate data (Shneiderman, 1996) and (Keim, 2002) which was *never used in forensic computing*. To present in an easily understandable way how parallel coordinate work, let authors consider an example. In Figure 7, the data contains 5 data points and each data point describes as a PC via 6 attributes (Identified data tampered features, Table 2) RecycleBin (RB), Internet History (IH), Cache (CSH), Cookies (COO), Restore Point (RP) and Prefetch (PR) can be seen in an  $A = 6$ -dimensional. The parallel coordinates map each dimension to a separate vertical axis (column).

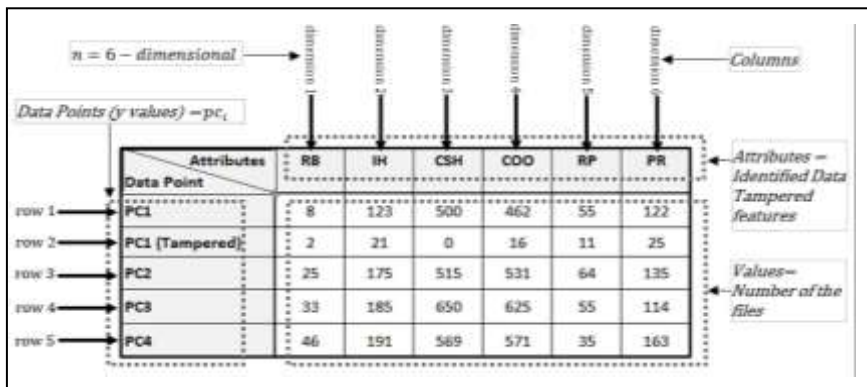


Figure 7: Designing visualization (multivariate dataset with several attributes or variables per data point)

However, instead of corresponding to the horizontal row, each data point  $pc_i$  is now mapped as a polyline that connect the points on the vertical axes whose coordinates (y values) equal the point attribute  $a_i$  (see Figure 7 and Figure 8).

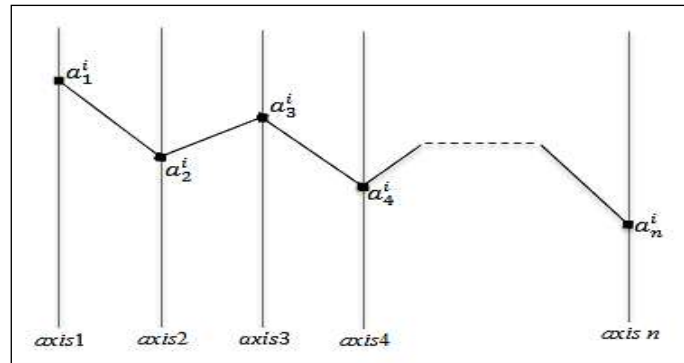


Figure 8: Designing visualization - parallel coordinates

2.2.2 Result (finding suitable visualization)

Following on the previous step, it is clear that what data, what size of data and what visualization type the authors are looking for. It was developed open-source software concerning parallel coordinate visualization, to import data (loading the file) and visualize them. Figure 9 illustrates the parallel coordinate visualization technique for the above dataset (see Table 3). There are six-axes; corresponding to the first six data attributes (data tampered features, Table 2). Each axis is scaled individually to show the full range of its attribute value (number of the file). Each polyline (row) represents a different PC of the determined five in the dataset. The polylines are drawn with a certain amount of transparency and areas covered by many lines. The red line and associated labels show the detail of the suspicious PC record under the mouse pointer (PC1 tampered). This visualization already shows several facts. A bunch of the lines that run parallel indicate a similar data point; the lines widely spread apart along an axis to show a large variation of the data attribute. It is apparent dataset distribution, the correlation between attributes (identified data tampered features), and data point (PC's) comparison.

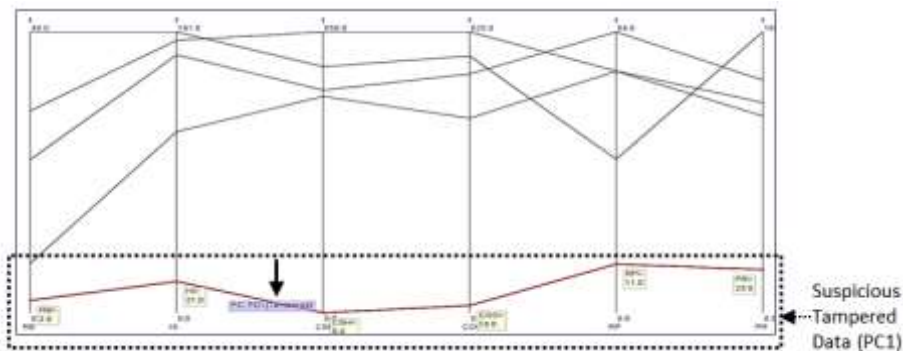


Figure 9: Parallel coordinates visualization view, scale axes 0-max (maximum number in each axis)

Other important features of this tool are viewing parallel coordinates visualization in two other scale axes types that present graph in an easily understandable way to the reader (see Figure 10 and Figure 11).

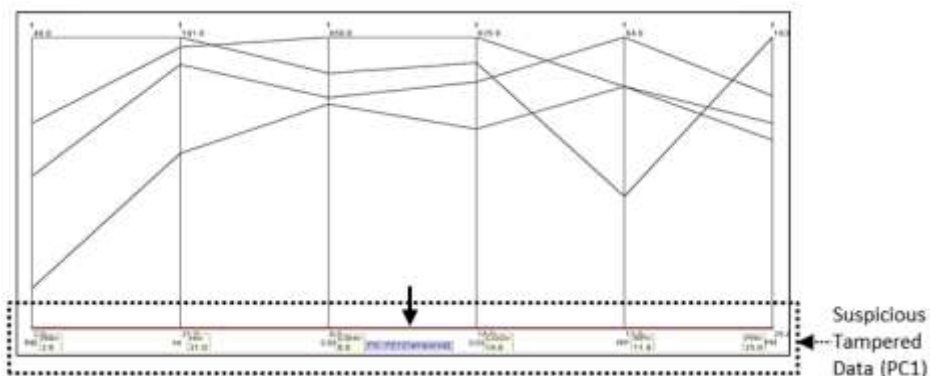
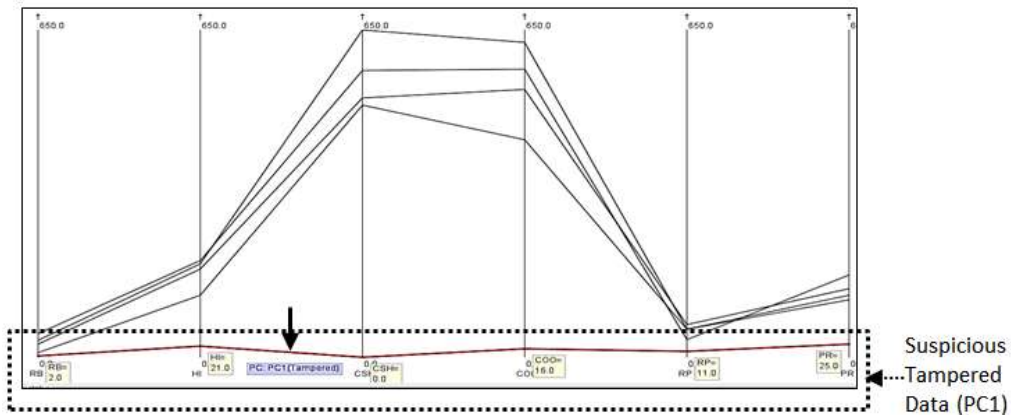


Figure 10: Parallel coordinates visualization view, scale axes min-max (minimum & maximum number in each axis)



**Figure 11:** Parallel coordinates visualization view, scale axes min-max (abs) – (minimum number in each axis and maximum number in all axis)

As presented above any action by a user on a computer, whether it's surfing the internet, communication or file storage, affects data stored in the computer. As a result, Parallel coordinates visualize PC1, PC2, PC3, PC4, and PC1 (tampered) behaviour in the normal way of usage in three different views. As explained earlier in step 2.1.3, I used the AF tool to wipe (or delete) data from PC1 for tampering purpose. The above figures show another PC's with some activities. But, PC1 has no activity and almost flat as displayed with a red line (indicated by arrows). PC1 seemed like a suspicious case of tampering for investigators just by first looking at the visualization. This is a remarkable result in the forensic field which never done before. When investigators facing with important items such as time and cost, there is no need to look for a needle in a haystack specific when an offender is familiar with the digital investigation process and no traces are left behind himself or herself. In this part, the authors identified suitable visualization to visualize identified data tampered features in a significant, meaningful and simple way to help the investigator to detect digital evidence tampering and data anomalies.

In summary, there are many situations when it is possible to determine whether the absence of evidence was caused by malicious tampering or normal system operation. To further reduce the chances of making a false-negative or false-positive conclusion, the investigator examining CADDED visualisations must take into account any prior information available to him/her that may suggest that misinterpretation is likely. In particular, the investigator must consult the system administrator of the organisation owning the investigated computer to establish the normal settings that may result in the natural absence of evidence (such as Recycle Bin settings, disk defragmentation settings or etc.) to avoid misinterpretation of CADDED visualisations.

Although it would be much preferable to establish some quantitative answers regarding the accuracy of CADDET at detecting evidence tampering, it was not practically possible during this study. Any attempt to do it would require access to evidential data from a large number of real cases with known instances of tampering and no tampering occurred. Unfortunately, the author had no access to such data, and the quantitative investigation of the error rate of CADDET method is left for future work.

### 3. Conclusion

The main goal of the current study is to identify tampered data features after applying any Anti-Forensics tools and visualize them in a comprehensive way for investigators. As a result, *computer-aided* diagnostic digital evidence tampering (CADDET) explored a semi-automated approach based on visualization of relevant data properties, helping human investigators to detect digital evidence tampering and anomaly.

One of the outstanding features of this research is its malleability. It can easily apply to parts or the entire dataset (depend on investigators' interest) in the digital devices, visualize, and reveal offender concealment behaviour concern to the detection of evidence tampering. It is clear that CADDET is capable of adapting to changes in technology and tampering behaviour over time, there is no risk when digital investigators are faced with an issue they have not encountered before. The explored method can apply to another operating system (for instance: Linux, iOS, Android, etc.) to detect digital evidence tampering and reveal offender concealment (or behaviour).

CADDET sheds new light and explores new insight to contribute investigators to fill a gap of detecting digital evidence tampering. This semi-automated approach had never been studied before and is novel in the context of digital forensics.

#### 4. Future work

Future work includes expanding the proposed approach to other operating systems such as Linux, IOS, Android, etc. to detect evidence tampering in the digital forensics field. We can also consider the importance of using the EnCase forensics tool, another important aspect of our future work includes the creation of EnScripts for Windows 7/8/10 and Windows Vista systems as well as adaptations for Linux and Mac operating systems, as well as EnScript functionality to parse SQLite databases and DAT files/registry hives to extract more information and add per-user artifacts analysis using substrings to determine usernames in location and analyse accordingly.

As well as for the visualization element, it would be much preferable to establish some quantitative answers regarding the accuracy of CADDET at detecting evidence tampering, however due to time limitation it was not practically possible during this project. Any attempt to do it would require access to evidential data from a large number of real cases with known instances of tampering and no tampering occurred. Unfortunately, the author had no access to such data, and the quantitative investigation of the error rate of CADDET method is left for future work. Besides, future work in the area of implementation of angular brushing is the next step in which we enable the space between the axes for brushing interaction. There are a few directions where work can be done to integrate this methodology in the field of digital forensics and evidence analysis. Hence, making CADDET acceptable in the court of law in the future.

The research reported in this paper is partially funded by the EU H2020 project AIDA (grant agreement ID: 883596)"

#### References

- AccessData [WWW Document], 2010. URL <https://www.pluralsight.com/courses/accessdata-forensic-toolkit-ftk-imager> (accessed 2.25.21).
- Acesoft, 2001. Tracks Eraser [WWW Document]. URL <http://www.acesoft.net/> (accessed 2.25.21).
- Bostock, M., 2020. d3/d3 [WWW Document]. GitHub. URL <https://github.com/d3/d3> (accessed 2.25.21).
- Chang, K., 2013. Parallel Coordinates [WWW Document]. URL <https://syntagmatic.github.io/parallel-coordinates/> (accessed 2.25.21).
- Davies, J., 2016. Parallel Coordinates [WWW Document]. URL <https://bl.ocks.org/jasondavies/1341281> (accessed 2.25.21).
- DFire, 2013. DigitalFIRE Labs [WWW Document]. URL [http://dfire.ucd.ie/?page\\_id=23](http://dfire.ucd.ie/?page_id=23) (accessed 2.25.21).
- Dfire, 2016. DFIRE Forensic Prolog [WWW Document]. URL <http://dfire.ucd.ie/?p=1478> (accessed 2.25.21).
- Dimmick, 2005. CCleaner [WWW Document]. URL <https://www.ccleaner.com/ccleaner> (accessed 2.25.21).
- GOG, S., 2002. Windows Eraser [WWW Document]. Download.com. URL [https://download.cnet.com/Windows-Eraser/3000-2144\\_4-10550310.html](https://download.cnet.com/Windows-Eraser/3000-2144_4-10550310.html) (accessed 2.25.21).
- Inselberg, A., 1997. Multidimensional detective, in: Proceedings of VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium. Presented at the VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium, IEEE Comput. Soc, Phoenix, AZ, USA, pp. 100–107.
- Keim, D.A., 2002. Information Visualization and Visual Data Mining [WWW Document]. <https://ieeexplore.ieee.org/>. URL <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=981847> (accessed 3.18.21).
- Shneiderman, B., 1996. The eyes have it: a task by data type taxonomy for information visualizations, in: Proceedings 1996 IEEE Symposium on Visual Languages. Presented at the Proceedings 1996 IEEE Symposium on Visual Languages, pp. 336–343.
- Tableau, 2015. parallel coordinates test - dmackay | Tableau Public [WWW Document]. URL <https://public.tableau.com/s/profile/dmackay#!/vizhome/parallelcoordinatetest/parallelcoordinatetest> (accessed 2.25.21).
- Tarau, P., 2014. Tarau Paul [WWW Document]. URL <http://www.cse.unt.edu/~tarau/> (accessed 2.25.21).
- Teelink, S., Erbacher, R.F., 2006. Commun. ACM 49, 71–75.
- Wikipedia, W., 1972. Prolog. Wikipedia. URL <https://en.wikipedia.org/w/index.php?title=Prolog&oldid=1000783550> (accessed 2.25.21)
- Wikipedia, W., 2016. Recycle.bin. Wikipedia. URL [https://en.wikipedia.org/w/index.php?title=Trash\\_\(computing\)&oldid=991812691](https://en.wikipedia.org/w/index.php?title=Trash_(computing)&oldid=991812691) (accessed 2.25.21)
- X-Ways Forensics [WWW Document], 2002. URL <https://x-ways.net/forensics/> (accessed 2.25.21).



# Weaknesses of IoT Devices in the Access Networks Used by People in Their Homes

Aarne Hummelholm

Faculty of Information Technology, University of Jyväskylä, Finland

[aarne.hummelholm@elisanet.fi](mailto:aarne.hummelholm@elisanet.fi)

DOI: 10.34190/EWS.21.036

**Abstract:** Today, the rapid development of information technology, components, systems, and applications poses major challenges for designers of networks, new services, and new types of smart devices to make devices and systems secure and usable in this digital world. Different types of smart devices are used everywhere, and people are being provided with more effective services to meet their everyday needs. Those smart devices are increasingly connected to many different types of sensors and IoT devices, whose security solutions are weak or non-existent due to the urgency of manufacturers to bring devices to market as quickly as possible. As a result, they have no time to create good security solutions for those devices. Those sensors, actuators and IoT devices are used in industrial environments, different types of municipal systems, different types of homes devices and systems, buildings' systems, free-time environments, healthcare systems, cars, ships, and so on. Access network devices are connected to smart devices or access nodes and through those access networks devices information is sent to data centers, where they use different types of services and store information they are collecting. People use their smart devices for different services and in different environments, and they often buy apps from app stores that may not be secure enough. Normal store-bought smart devices do not include hardened functionality and are therefore easy to hack or malware can easily be installed on them. Those vulnerabilities give hackers and cyber attackers possibilities to attack systems via those sensors, actuators and IoT devices and install malware in those systems and on devices. Due to the rapid development of sensors, actuators, and IoT devices, hackers and network attackers are aware of this and are developing new types of malware that are optimized for use in such environments. One example of the latest Malware is Mirai, which has been used in attacks against smart devices and processor controllers, IoT devices, actuators, and sensors. It is also used against routers, gateways and switches. There are also new variants that may be even more harmful when they are used against systems. Even if we make the communication connection from smart devices to data center secure enough, it is still not enough because the sensors, actuators and IoT devices are full of vulnerabilities.

**Keywords:** sensors, actuators, IoT devices, vulnerabilities, security, attacks

---

## 1. Introduction

Digitalization changes societies, especially the ubiquitous use of information and communication technology (ICT) in daily services, societal services, health and wellbeing and others. Citizens must be able to use those services wherever and whenever they need them and in real time. This means that ICT systems must be connected to each other because citizens use their smart devices everywhere when they use these digitalized smart society services. These devices may not have security systems to protect them from hackers and network attackers. In health and wellness (H&W) services, which are some of the most critical services in society, security and cybersecurity issues are one of the most important issues in this digitalized environment. The rapid development of information and communication technologies makes some of these issues even more urgent, making cybersecurity an important ethical dimension of future H&W solutions (Christen, Gordijn and Loi, 2020; Weber and Kleine, 2020). These cybersecurity issues are becoming increasingly important as citizens use the services of smart societies from different access networks and from their homes, and perhaps their home environments are not protected to a high enough level when, for example, we consider cybersecurity attacks against H&W services. Today these services are often used at home in home access networks, which are connected to the Internet or mobile networks, meaning that home networks are a critical environment now and also in the future.

Most IoT devices, sensors, and actuators use the same frequencies and frequency bands, which means that there is a lot of interference in such environments. Such devices are found in buildings, offices, industry, cars, ships, trains, airplanes, and so on. In short, everywhere people usually move. A number of investigative reports by journalists have presented examples of how cyber-attackers and hackers can strike against information systems in office buildings, hospitals and peoples' homes, among other environments. The latest malware, Mirai and its variants, are now being used against many systems, including those in homes. These include attacks against home routers (ENISA, 2016) (Pen Test Partners). Mirai is a piece of software that is used to form a malicious botnet, a network of connected devices (bots) that can be controlled to attack others on the Internet. This is done without the owner's consent or knowledge. Generally, these attacks take the form of distributed denial of

service (DDoS) attacks. This involves hundreds or even thousands of bots sending traffic to a server, consuming resources, and stopping the server from responding (Pen Test Partners). As people use the latest possible Internet-connected devices in their homes and strive to make them smarter and more efficient, they need to know how to protect those connected devices.

Such a rapidly evolving digital operating environment requires entirely new perspectives, approaches, and research methods to identify cybersecurity threats and risks in the devices, systems, and application levels, end-to-end, including telecom and service operators, to make systems secure to use. This study not only analyzes devices, systems, applications, or systems in silos, but takes into account the entire operating environment, with its dependencies, threats, and risks that can be due to many different factors in the current digitalized and interconnected world. Currently, most studies examine only one system or device and do not look at all the operating environments involved in it. We must also consider the threats posed by cyberwarfare. In this kind of rapidly evolving digital operating environment, we need to make these analyses at the level of systems and environments that form the entire system together, not just on one device or one system level. This is a new way to conduct these studies. Due to the limited number of pages, this research has been limited to access networks and their systems, as well as how IoT devices, actuators, and sensors are connected to smartphones through access networks and how their data travels from those smart devices to telecommunication networks edge systems and data centers. The one research question is as follows: How can we prevent or reduce radio frequency interference and eavesdropping on access networks and thus also reduce the possibilities for cyber-attacks?

## 2. Structure of this paper

The purpose of the study is to find architectures and implementation models for secure network environments at home so that people there can safely use the services provided to them by society in these interconnected environments. Section 3 describes the home environment of the future, section 4 discusses the telecommunications solutions in the operating environment, and section 5 discusses the cybersecurity threats and challenges related to the operating environment and presents solutions to improve the situation. Section 6 presents some conclusion and proposes directions for future study.

## 3. The home environment

Applications of intelligent sensor systems, analytical devices, telecommunication systems, and various information systems are rapidly evolving and adapting to people’s daily activities and needs, including more efficient and intelligent home-related services. Home environment systems include many services people using every day in their homes. Figure 1 shows smart society services people are using from their homes with their smart devices and through their home access networks every day. When those information services are interconnected, they form complex environments. When people use them with their smart devices, the security of these devices becomes critical, particularly if the devices lack security software or protection systems against malware or its variants. The security gaps in these devices may be one route to infecting society’s services with malware.



Figure 1: Services of the smart society used by citizens from their homes in the future

If we think about health care systems in this kind of environment, this development means that in the digital world, people, including older people, can be provided with more effective treatment methods that will allow them to live longer and better in their homes. Better home care and preventive healthcare will become more



available. People can easily carry portable sensors and intelligent devices in their bodies and wrists that relay their vital information to hospital systems in real time, which can be monitored by healthcare staff. Peoples can also easily follow their own vital signs in real time from their smart devices. Along with healthcare services, people use a wide variety of IoT devices, smart sensors, and actuators that use the same frequencies as the example home sensors use. This overlap presents a number of threats and creates significant challenges for cybersecurity. Figure 2 shows smart society services in different segments and services inside segments that people use every day from their homes.

The left side of Figure 2 presents scenarios, use cases, requirements, dependencies, risks, and threats. Then there are the architectures, national and international guidelines, regulations, directives, and specifications. All of the smart society’s ICT systems (ICT) are regulated at the systems and applications level in many ways and there are many recommendations in use. The devices in use are also regulated in many ways. There are also recommendations, regulations, and guidelines on security and cybersecurity issues in those environments. The devices used by people are freely available for purchase from stores, which they then use with these services. The software that comes with these devices may not have been tested and analyzed. The devices used may lack security software and cybersecurity-related software, and so on.

Looking at Figure 2, we see that these questions represent an enormous challenge from the perspective of the services and information systems of smart societies. We can also see the ways in which hackers and cyber-attackers can attack societies’ information systems. In many countries they have done just that. One of the most critical systems in these cases have been the healthcare systems in the hospitals, which are also used from homes. People use healthcare services with the devices they have directly purchased in stores. This type of situation means there are many attack vectors, and it is almost impossible to find every one and identify all the possible vulnerabilities cyber-attackers and hackers can use against information systems.

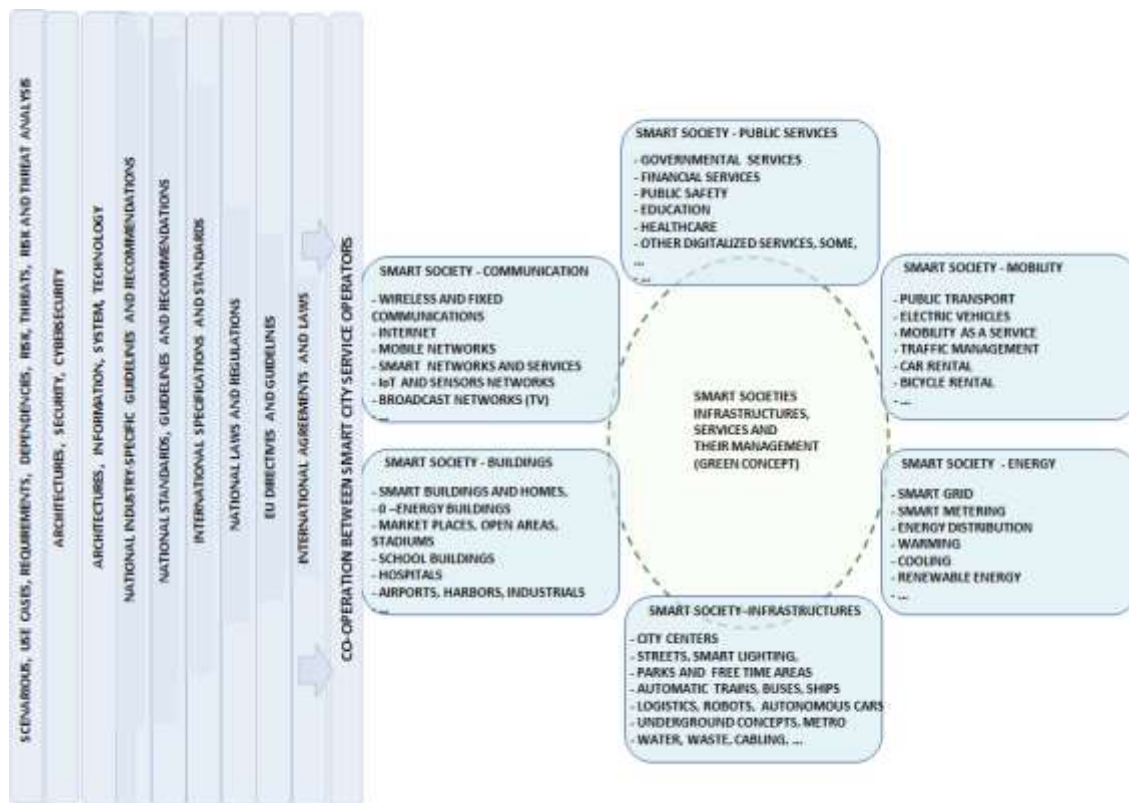
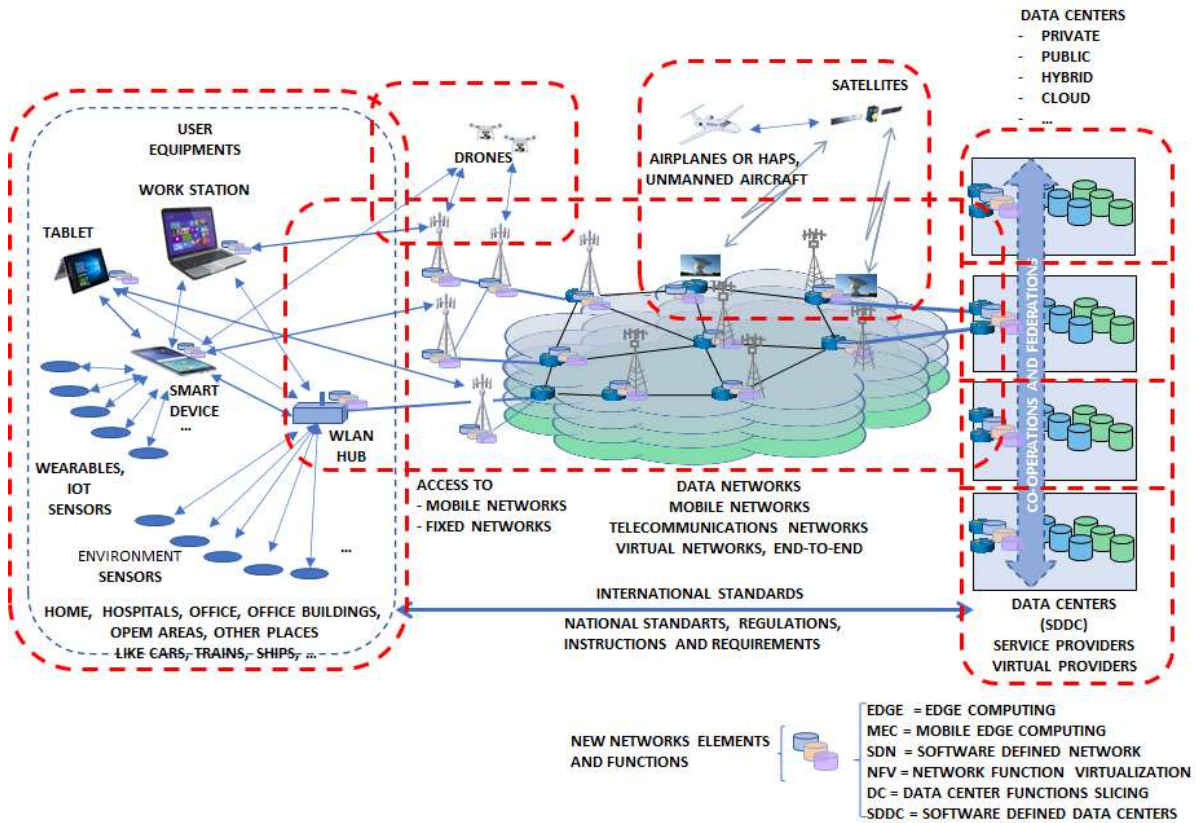


Figure 2: The service segments of the smart society with requirements and regulations (Hummelholm, 2019, p. 30, modified)

#### 4. The communications environment and systems

Nowadays, people can be in shops, homes, leisure places, theaters, and anywhere else they need to go to do the necessary things and they have their smart devices with them. When we use new types of smart devices and services, we also need to verify, in one way or another, that telecommunications operator networks and services

in data centers are provided according to requirements. This is so that we can be sure that service systems work in a secure and trustworthy way, and that privacy and security issues are safely handled (EU-NIS, EU-GPDR and EU-MDR). Figure 3 presents the communications technical environment of smart societies and the systems used there. Those communications and service environments are virtualized in the future from end to end. Figure 3 shows how complex smart societies information and communications systems will be in the future. These systems will require integration between different information systems, collaboration, orchestration, and cooperation so that this smart connected information society works well in every situation. This must be considered when access networks and the systems used in them are being developed (see Figure 4).



**Figure 3:** Communications technical environment and systems (Hummelholm, 2018, pp. 523–532, modified)

Figure 4 shows the home environment’s sensors and devices, which often use the same frequencies. This is a significant problem in home environments because e-health and m-health sensors of patients’ and elderly persons’ use the same frequencies, and they are connected to the patients’ and elderly persons’ smart devices. This means possible interference between different devices so that those healthcare systems may not work at all. Some of these devices use information that has different levels of security, such as public, restricted, or confidential information. Today, there are many vulnerabilities in these IoT devices, sensors, and actuators, so cyber-attackers and hackers can attack systems in such environments. These attacks are already carried out in the world today. Here are a few examples of those attacks that have been reported on by the worldwide media. In recent years, the news media has reported on a range of attacks in various sectors, including on medical devices (Palmer, 2018), the healthcare sector (Davis 2019, WinWire year, Westman 2019), home routers (ENISA, 2016), and 5G networks (Borgaonkar, 2019).

In Figure 4, an MEC router and EDGE router are outside of the home and they are installed and managed by network operators. In Figure 5, we see examples of smart device connections to the smart society services when as they are used them from a home network environment. They can add a level of convenience to people’s lives, but they could also make home and connected devices vulnerable because they are also connections to the Internet and other networks in a smart society.

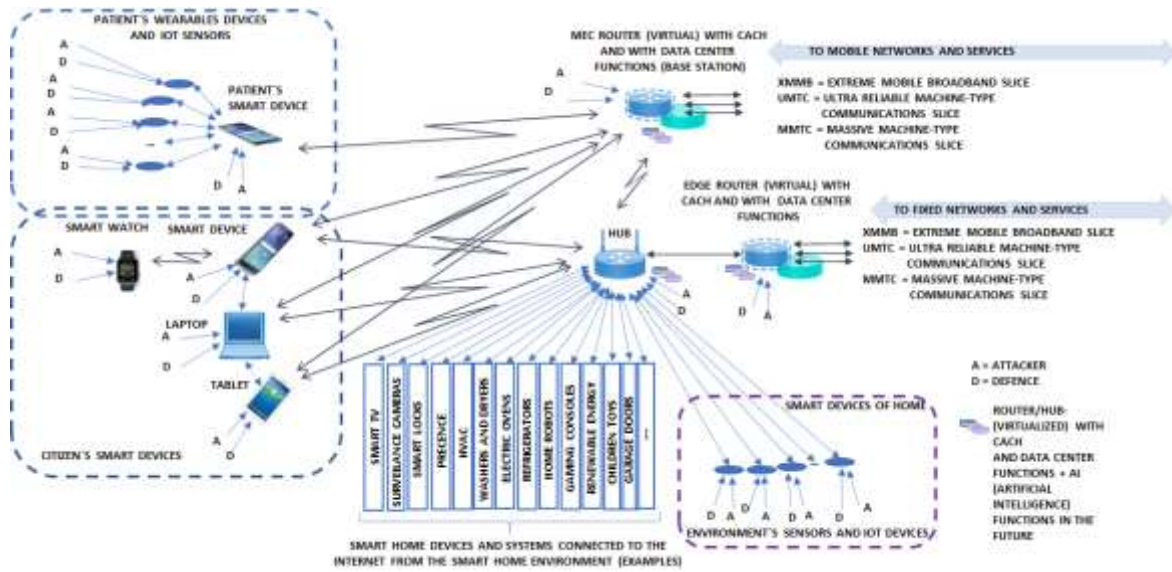


Figure 4: Home area network in smart homes and connections to smart devices

### 5. Cyber threats against home access networks and systems

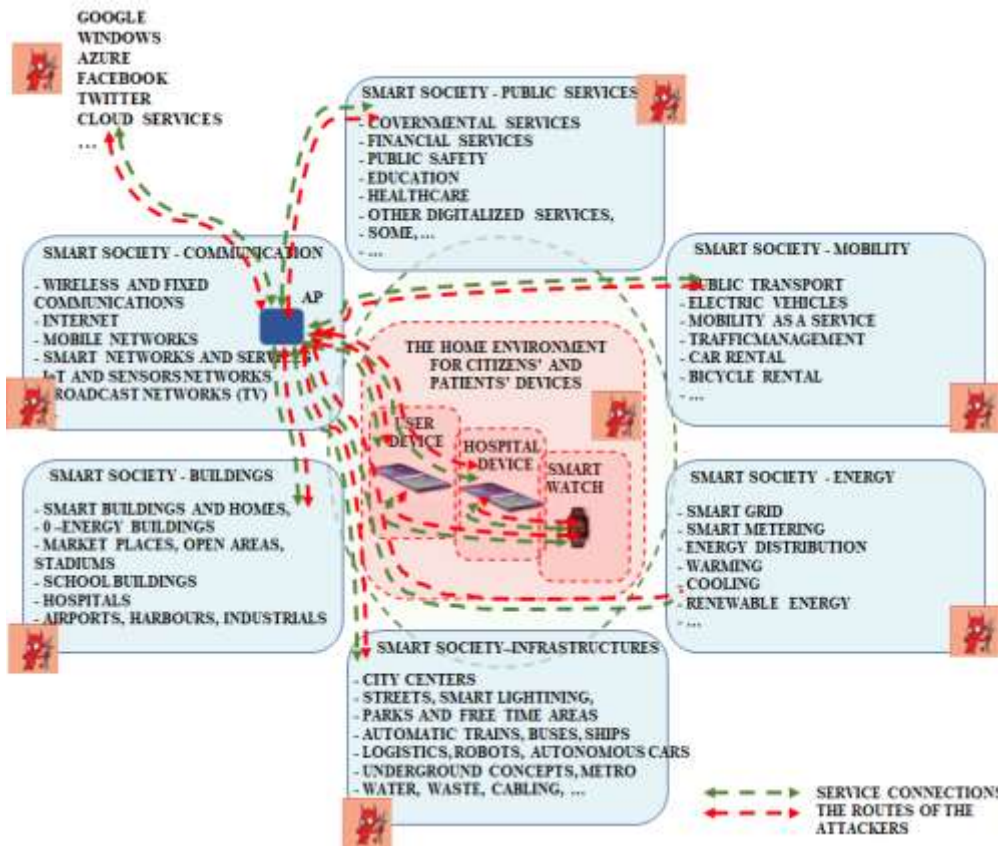


Figure 5: Smart device connections from home environment to smart society services

When people use smart services through smart devices, they may not be aware of the information security or cybersecurity of those services in those different segments (see figures 2 and 5). Because today almost all systems are interconnected in one way or another, and if a single smart device or network-connected sensor contains vulnerabilities, hackers and cyber-attackers can find those vulnerabilities and exploit them quickly. Users may not notice the situation at all and continue to use the services of society even when their devices are compromised. From Figure 5, it is difficult to find and define an end-to-end service chain and ensure security issues because many of the service operators' services nowadays use cloud-environments. Even if users have cryptographic systems on their smart devices from home to the data center end to end, that is not enough,



because IoT sensors or other sensors connected to smart devices often lack any kind of security system. Figure 6 shows a smart device and the associated wellness devices, medical devices and home devices that use the same frequencies and maybe include no encryption system. These connections are open to hackers and cyber-attackers and these systems work in home access networks (see Figure 4). Table 1 shows the vulnerabilities of access network sensors and IoT devices.

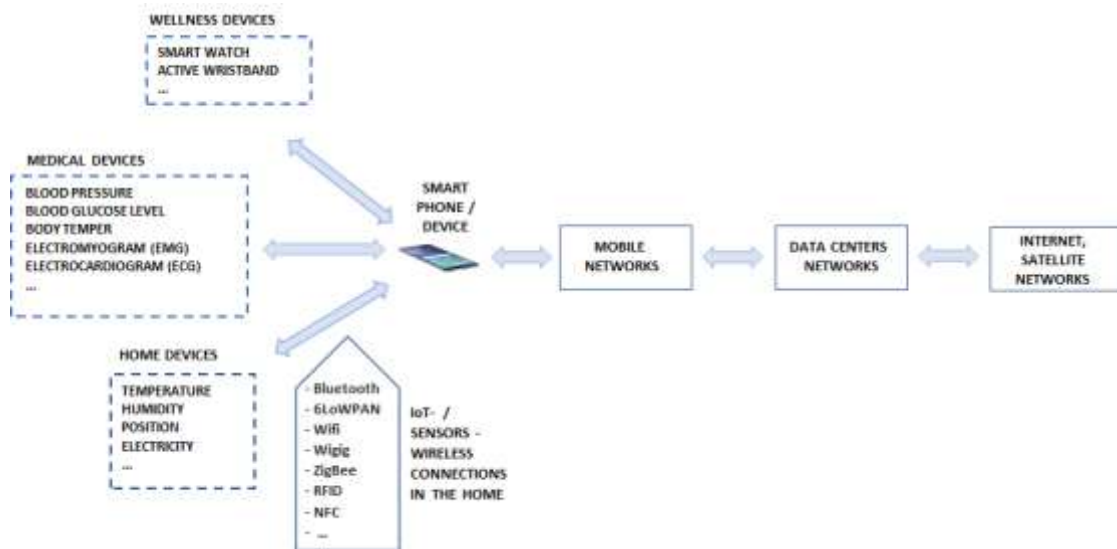


Figure 6: Smart devices in the home environment

Table 1: Vulnerabilities of access network sensors

TYPE	DEVICES (FROM FIG. 12)	TYPES OF COMMUNICATION SYSTEM	EXISTING CONTROL IN ACCESS SIDE	EXISTING CONTROL IN DATA CENTER	VULNERABILITIES IN IOT-/SENSOR-DEVICES	VULNERABILITIES IN ACCESS CONNECTION	ATTACK PROBABILITIES
DEVICES							
MEDICAL DEVICES	WIRELESS CONNECTED DEVICES (SENSORS AND IOT DEVICES)	BAN, PAN, WAN	ID CODE, SECURITY VPN OR NONE	FIREWALL	OFTEN NON-SECURITY SOLUTION	VULNERABLE WIRELESS CONNECTION	VERY HIGH
WELLNESS DEVICES	WIRELESS CONNECTED DEVICES (SENSORS AND IOT DEVICES)	BAN, PAN, WAN LTE, 3G, 4G, 5G, 6G	ID CODE, SECURITY VPN OR NONE	FIREWALL	OFTEN NON-SECURITY SOLUTION	VULNERABLE WIRELESS CONNECTION	VERY HIGH
PEOPLES HOME ENVIRONMENTS' DEVICES	WIRELESS CONNECTED DEVICES (SENSORS AND IOT DEVICES)	PAN, WAN	ID CODE, SECURITY VPN OR NONE	LOCAL OR REMOTE, (FIREWALL)	OFTEN NON-SECURITY SOLUTION	VULNERABLE WIRELESS CONNECTION	VERY HIGH

WBAN = WIRELESS BODY AREA NETWORK (~ 1 – 2m)  
 WPAN = WIRELESS PERSONAL AREA NETWORK (~10 – 100 M)  
 WLAN = WIRELESS LOCAL AREA NETWORK (~50 – 100M AND WITH CERTAIN SOLUTIONS FURTHER)  
 LTE, 3G, 4G, 5G, 6G = MOBILE NETWORK TECHNOLOGIES

WE ALSO NEED TO TAKE CARE OF  
 - PRIVACY ISSUES  
 - ETHICS AND MORLAS ISSUES

Table 1 shows that even when a smart device has a good encryption system, the sensor side and sensor connections to those devices are not protected enough or at all in access networks. These wireless connections between IoT devices and sensors provide the ability to listen to radio frequency signals and perhaps take critical information from devices and put malware on those devices. One example of the latest malware is Mirai, which is used to attack smart devices and processor controllers, IoT devices, actuators, and sensors, and is also used against home systems. It is used against routers, gateways, and switches. There are also new variants of Mirai and they may be even more harmful when used against systems. In turn, these kind of home devices can be easily exploited for eavesdropping on in home environments as well (see Table 1).

### 5.1 Making and modelling of threat analyses and situation awareness

Chapter 4 shows that access networks have many problems when they use wireless technology to connect IoT devices and sensors. We can say that all such connections are vulnerable and can be used by attackers for their own purposes (see Table 1). To threat analysis the devices in an access network, we can use an attack-tree model (Figure 7), based on figure 4. From figures 2–4 we can determine the service chains, which system or service is

used at any given time, which services the smart devices use inside the home when connected when connected to services, what information is retrieved, and so on. Risk and threat analysis as well as probability calculations must be done for every sensor, for every smart device, GW devices to networks outside of home, applications, software, systems, used protocols and data models in every service. When we have analyzed and calculated risks, threats, and probabilities, we can put those values into Table 3, Threats and Risks Table.

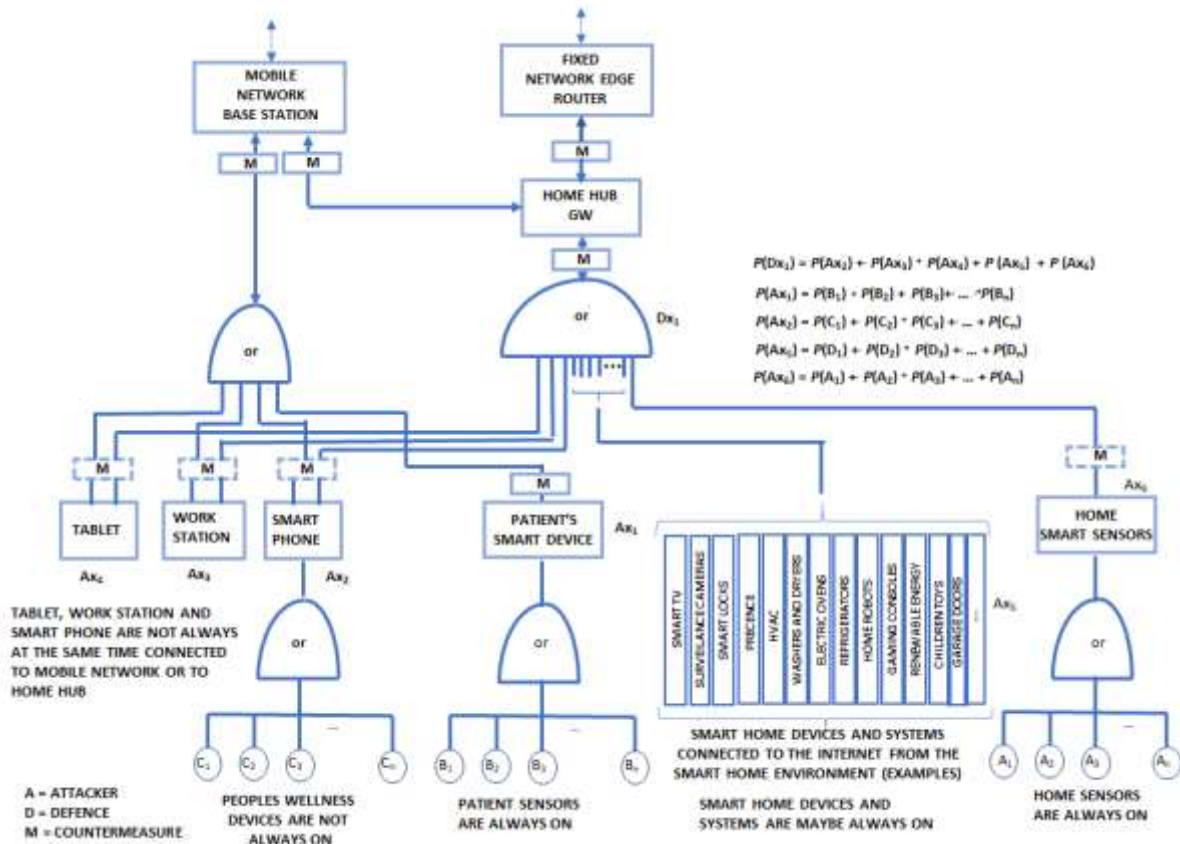


Figure 7: Attack tree model for access network based on Figures 3–6

The probabilistic success of attacks ( $P(t)$ ) against a device  $x$  in a home can be calculated as follows (Wang and Liu, 2014)

$$P_{Ax}(t) = p_A(1 - p_D(t))(1 - p_M(t)) \tag{1}$$

Probability Calculation of Successful Attacks: -  $P(Ax_1)$  occurs if  $B_1$  or  $B_2$  or  $B_3 \dots B_n$  occurs, -  $P(Ax_2)$  occurs if  $C_1$  or  $C_2$  occurs or  $C_3 \dots C_n$  occurs, -  $P(Ax_6)$  occurs if  $A_1$  or  $A_2$  or  $A_3 \dots A_n$  occurs. If we look at sensors and smart devices related to the home network environment, in figures 4 and 6, smart devices or home access networks may not have security or encryption systems that could be used to protect systems from hackers and cyber-attackers. The same situation can exist anywhere in the access network systems of smart societies. In the same way, the threats and risk probabilities of all the devices connected to home networks can be calculated.

Table 2: Meaning of notations

ACTION	EXAMPLES	NOTATION
ATTACK	SNIFFING, ENUMERATION, SCANNING, ...,	A
DETECTION	PORT SCAN, INFORMATION SCAN, ...,	D
COUNTERMEASURE	ANALYSING OF VULNERABILITIES, SAFEGUARDS PUT IN PLACE, ...,	M

**Table 3:** Model for a threats and risks table (Hummelholm 2019, pp. 641–649)

Ref ID	Org	Functions	Category	Threat	Threat/ Risk	Existing Control	Threat /risk level			Accept /reduce	Recom- mented control	Residual Threat/ Risk			Check Point
							L	C	R			L	C	R	
1 / AH	MC	Identify	Access	Foot - print ing	Target Access	IDS /IPS	3	3	8	Reduce	EU- dir.	2	2	3	xx

Where L = Likelihood, C = Consequence, R = Risk

## 6. Conclusions and future work

Analyzing the architectures of smart home infrastructures and services while considering the needs of all was a challenging task because smart homes are large and complex environments and structures. In these environments, people need to use services securely and in real time. That is why I have distributed smart cities’ services to different segments of the infrastructure on the basis of different functions (see Figure 2), so that we could get a better picture of the service environments we have ahead of us. It was also difficult work to access the technical specifications of IoT devices, sensors, and actuators. Technical information is needed to check the encryption solutions of devices and to see which data centers information from devices would be going to. In many areas, development work occurs in their own silos, and segment developers often do not talk to each other. It was difficult to obtain sufficient information on, for example, communication and information systems inside buildings, their development plans, internal networks in buildings, and security and safety mechanisms in buildings.

It is difficult to carry out analyses and revisions to take better account of the new security requirements and cyber security directives. These difficulties lead to sub-optimizations and not to the optimization of the development of information systems for smart homes and smart societies. Information from security and safety mechanisms in office buildings is important because we can use the same technologies in home access networks. There are many IoT devices, sensors or actuators being introduced into systems, and many smart devices are being used. These may be full of vulnerabilities and shortcomings in security mechanisms, and their security solutions are not good enough. Hackers and cyber-attackers can attack the service system of smart homes through those devices and sensors, even if the attackers are on different continents using satellite connections or submarine cable connections. When people use smart devices for e-health purposes, security issues are even more critical in this kind of communications environment. These problems were the reason why cybersecurity analyses were not performed on only one device, system, or environment in their own silos, but analyses were performed on various systems, taking into account all dependencies on operating environments, systems and devices in use. This also required that we do a lot of enterprise and security architecture work.

We have an ongoing telemedicine project, and in which we are looking for a model and solution to protect patient devices and residential communications networks so that patients can use telemedicine devices safely at home. At the same time, we are also looking at the devices of all home users and the protection of their devices in such an environment (ENISA, 2015).

A few issues can already be raised on how to better protect home networks, users’ devices, and home network connections to telecommunications networks and the services provided through them. These issues include the following:

- New frequency allocation for IoT devices, sensors and actuators depending on use (new microwave bands).
- In the Home HUB, enable the separation of services: health care, welfare, banking, taxation, school, etc.
- Utilization of artificial intelligence in the interfaces of devices and networks.
- Separate equipment with encryption systems for patients (operator provides equipment management and defines and implements security solutions)

- Dividing operating environments into security zones according to security needs (e.g., home, hospitals, office)
- Using VPNs from a smart device to services (mobile phone, workstation, tablet).
- A new generation of devices on which it is possible to use new types of encryption systems.

## **7. Future work**

Home network security issues have not received much attention in the past, but now the situation is changing. As a result of this new rapid technological development, more attention needs to be paid to home information networks and cybersecurity in home networks. Solutions are needed to ensure the safe use of peoples' smart devices when accessing societal and other services.

We need a clear picture of these environments in smart home environments, because, for example, threat analyses are difficult to do. As more and more IoT and sensor devices are connected to smart home telecommunication networks, we know little about these.

Technical progress is accelerating rapidly and the importance of analyzing individual components needs to be considered openly. Even though dependencies, dependency analyses, threat assessments, and threat analyses could be made comprehensively in some areas, it is essential that threat assessments and analysis of future societal services are made with the support of computer programs as well as of artificial intelligence.

We must develop and test a high-level architecture description model, such as the enterprise architecture framework and target architecture models (DRAGON1-open) (JHS), so that we can better describe the infrastructures and services of smart homes and buildings.

When the vulnerabilities of different OSI layers, and the various vulnerabilities associated with the protocols in use are added to this whole, the number of issues to be verified increases. The inherent defects and vulnerabilities of encryption solutions and cryptographic network solutions must also be analyzed.

The functionality of the virtual access network services and the terminals which use them must be one research area in various attack situations. It would be important to study, for example, remote health care systems, rescue authorities' systems, security authorities' systems, and other critical systems.

The use of artificial intelligence needs to be studied and tested for its ability to protect IoT devices, sensors and other e-health systems so that we can better protect these devices from malware and cyber-attacks and also monitor data flows from sensors to data center storage systems.

One research topic is to measure and test various frequency-induced interference methods in smart homes, office buildings, hospitals and other environments. One goal would be to check whether they influence people's devices and the services they use. Radio frequencies also affect the body in real time, so radiation levels must also be analyzed, and an attempt made to find the right level so that the patient is not harmed by this radiation.

Energy efficiencies are another important area to investigate in the smart home environment, communications systems, smart devices, and smart services from communications environments.

The use of artificial intelligence in analysis should also be examined, because its use would speed up work and provide opportunities to quickly identify vulnerabilities and fix them. As a result, security measures could be quickly targeted at the right location to prevent potential penetration into multiple networks and services, thereby preventing, as much as possible, the effects of attacks in virtual environments. A further topic could be the use of artificial intelligence for real-time tracking and analysis at different gateway (GW) points to analyse network threats and security and systems management systems in the smart cities environments.

One new area of research would be a quantum encryption system for use in wireless networks (4G, 5G, 6G, 7G) and services in the future. A further research area would be to find new frequency bands for the use of IoT sensors in healthcare services in microwave bands, so they cannot be hacked, eavesdropped on or otherwise inferred with so easily.

As we spend millions of euros on the demands of hackers and cyber attackers, it would be worth it to develop and establish different security zones for hospitals, homes or other critical environments to protect sensitive systems from radio interference and eavesdropping. After all, radio frequency interference can prevent an entire operating environment from working without any physical connection to any part of the network, depending on the location, of course. But if there are no security zones, it will be fairly easy for hackers and cyber attackers to attack against our systems.

## References

- Borgaonkar Ravishankar, "New vulnerabilities in 5G Security Architecture & Countermeasures", SINTEF INFOSEC Cybersecurity Research Group, 8. August 2019, <https://infosec.sintef.no/en/informasjonsikkerhet/2019/08/new-vulnerabilities-in-5g-security-architecture-countermeasures/>
- Christen, M., Gordijn, B. and Loi, M., "Introduction, in The Ethics of Cybersecurity", Cham, Springer, 2020, pp. 1-8.
- Davis Jessica, "Hackers Targeting Healthcare with Financially Motivated Cyberattacks", August 21, 2019, HEALTH IT SECURITY, xtelligentHEALTHCARE MEDIA, <https://healthitsecurity.com/news/hackers-targeting-healthcare-with-financially-motivated-cyberattacks>
- Dragon1-open, EA Method / Visualization Standard, 'Enterprise Architecture Framework', <http://wigi.dragon1.org>.
- ENISA, Security and Resilience of Smart Home Environments, Good practices and recommendations, DECEMBER 2015, [www.enisa.europa.eu](http://www.enisa.europa.eu)
- ENISA, "Mirai" malware, attacks Home Routers, December 14, 2016, <https://www.enisa.europa.eu/publications/info-notes/mirai-malware-attacks-home-routers/>
- EU-GDPR, The General Data Protection Regulation, 2016/679.
- EU- MDR, The Medical Devices Regulation, 5/2017.
- EU- NIS, 'Concerning measures for a high common level of security of network and information systems across the Union', 6/2016.
- Hummelholm, A., "Cyber threat analysis in Smart City environments", 17th European Conference on Cyber Warfare and Security, pp. 523-532, 2018.
- Hummelholm, A., "Cyber Security and Infrastructures of Smart Societies", 2019, p. 30.
- Hummelholm, A., "E-health systems in digital environments," 18th European Conference on Cyber Warfare and Security, pp. 641-649, 2019.
- JHS 179, 'Enterprise architecture planning', Modified date 2018-01-30, <http://www.jhs-suositukset.fi/web/guest/jhs/recommendations/179>.
- Kumar Mohit, "New Attacks Against 4G, 5G Mobile Networks Re-Enable IMSI Catchers", The Hacker News, February 25, 2019, <https://thehackernews.com/2019/02/location-tracking-imsi-catchers.html>
- Louis Scialabba, "IoT, 5G Networks and Cybersecurity: Safeguarding 5G Networks with Automation and AI", September 18, 2018, <https://blog.radware.com/security/2018/09/iot-5g-networks-and-cybersecurity-safeguarding-5g-networks-with-automation-and-ai/>
- Norton, 12 tips to help secure your smart home and IoT devices, Aug. 28, 2019, <https://us.norton.com/internetsecurity-iot-smart-home-security-core.html>
- Palmer Danny, "IoT security warning: Cyber-attacks on medical devices could put patients at risk", March 15, 2018, ZDNet, <https://www.zdnet.com/article/iot-security-warning-cyber-attacks-on-medical-devices-could-put-patients-at-risk/>
- Pen Test Partners LLP, Security consulting and testing services, "What is Mirai? The malware explained", <https://www.pentestparnres.com/security-plock/what-is-mirai-the-malware-explained/>
- Wang, P. and Liu, J.C., 'Threat Analysis of Cyber- attacks with Attack Tree +', 2014.
- Weber, K. and Kleine, N., "Cybersecurity in Health Care," in The Ethics of Cybersecurity, The International Library of Ethics, Law and Technology 21, Cham, Springer, 2020, pp. 139-156.
- Wetsman Nicole, "Health care's huge cybersecurity problem", Apr 4, 2019, The VERGE, <https://www.theverge.com/2019/4/4/18293817/cybersecurity-hospitals-health-care-scan-simulation>
- WinWire," Importance of Cybersecurity in the Healthcare Industry", WinWire Technologies, June 1, 2020, <https://www.winwire.com/healthcare-cybersecurity/>
- Zorz Zeljka, "New privacy-breaking attacks against phones on 4G and 5G cellular networks", Help Net Security February 25, 2019, <https://www.helpnetsecurity.com/2019/02/25/privacy-attacks-4g-5g-cellular-networks/>



# Cyber Security Analysis for Ships in Remote Pilotage Environment

Aarne Hummelholm, Jouni Pöyhönen, Tiina Kovanen and Martti Lehto

Faculty of Information Technology, University of Jyväskylä, Finland

[aarne.hummelholm@elisanet.fi](mailto:aarne.hummelholm@elisanet.fi)

[jouni.a.poyhonen@jyu.fi](mailto:jouni.a.poyhonen@jyu.fi)

[tiina.r.j.kovanen@jyu.fi](mailto:tiina.r.j.kovanen@jyu.fi)

[martti.j.lehto@jyu.fi](mailto:martti.j.lehto@jyu.fi)

DOI: 10.34190/EWS.21.025

**Abstract:** International and national maritime transportation systems are essential parts of critical infrastructures in every society. Digitalization makes possible to increase levels of autonomy in maritime transportation systems. In the research point of view, it will be done step by step. In Finland, to develop the remote pilotage on fairway environment is an example of this process. The function of all legacy and modern ships are essential parts of its cyber security. The integration of operational systems and information systems on ships take place on the ship's intranet in order to improve the cooperation of the various functions of the ships. Integration brings together new services and functions coming from digitalisation, the development of new maritime and autonomous traffic managements, the integration of monitoring and control functions, and the development of functions streamlining services. As ships increasingly use information systems to exchange information between different integrated systems on ships and ground systems, and between on-board operating systems and on-board communication systems, the examination of the digital structure and its dependencies is a very important part of cyber security analysis work. It enables to identify different functions in the system level, carry out risk assessments and identify their residual risks with sufficient accuracy. In the same way, the dependencies of different information systems need to be considered and, based on these dependencies, security and cyber security risks need to be identified. This paper presents a probability approach to cyberattacks versus a probability to defend attacks and at the end to evaluate cyber security risks related to the operations of ships. Cyber security of information systems used on board ships must be approved before the systems are put into service. In order to define this large entity in a controlled way into different parts, one good way is to use the enterprise architecture framework and its methods to visualize entire systems, subassemblies, and dependencies. The paper examines these ship's systems cyber security elements and the necessary security mechanisms to better manage the information and cyber security situation awareness now and in the future integrated operating environments on the way to autonomous ships.

**Keywords:** critical infrastructure, remote pilotage, ship, cyber security risks, probability, situation awareness

---

## 1. Introduction

ENISA emphasizes maritime transport as a crucial activity for the European Union economy. It enables large scale of trade, import and exports of goods and transport of passengers. In the European Union alone, maritime transport comprises around 1,200 ports. The global digitalization trend has led authorities to set recent policies and regulations to maritime transport stakeholders to face new challenges with regards to information and communication technology (ICT). This means new challenges in cybersecurity that covers as well as the Information Technologies (IT) than Operation Technologies (OT). (ENISA, 2019)

On the other hand, digitalization makes possible to increase levels of autonomy in maritime transportation systems. At the same time international and national maritime transportation systems are parts of critical infrastructures in every society. Challenges in cybersecurity are also relevant for critical infrastructure and its essential services point of view. There are many security threats. The latest challenges for the operating environment come from e.g. heterogeneous telecommunication networks where new devices and systems are seamlessly interconnected. These systems are now coming more intensely into new smart solutions like buildings, hospitals, ships and ships terminals and other transportation systems.

There are many research projects under work related to automated maritime transportation systems. In order to develop maritime autonomy in the first stage in Finland, the Sea4Value / Fairway (S4VF) research program has been established to create automated remote pilotage fairway features (ePilotage) in the ongoing research period from the beginning of 2020 to the end of January 2022. The S4VF program plan concerns a smart maritime transport system and research on automated remote pilotage fairway navigation. This smart fairway navigation ensures a channel which the existing vessels and future autonomous ships can use to travel safely and in the transfer of goods. The present situation recognition of fairway navigation is important in piloted waters and their approaches. In this context, the term "piloted waters environment" includes general fairway areas, often

in archipelagos, where navigation has been considered so difficult that additional, specialized competence is required from the navigator. Currently, this pilotage can either be performed by a pilot or by a qualified ship's master or officer holding a pilot exemption certificate. (DIMECC Oy, 2020)

International Maritime Organization, IMO (2017) paper "GUIDELINES ON MARITIME CYBER RISK MANAGEMENT" provide high-level recommendations for maritime cyber risk management. According to these Guidelines, "maritime cyber risk refers to a measure of the extent to which a technology asset is threatened by a potential circumstance or event, which may result in shipping-related operational, safety or security failures as a consequence of information or systems being corrupted, lost or compromised". In that sense all stakeholders are recommended to take necessary steps in order to add cyber risks to the overall shipping risk process. It means cyber risks "from current and emerging threats and vulnerabilities related to digitization, integration and automation of processes and systems in shipping". (IMO, 2017)

The paper of "The Guidelines on Cyber Security Onboard Ships Development" (2020) includes guidelines for implementation and maintenance of a cyber security management program for ships. In the paper is mentioned: that "it is important that senior management stays engaged throughout the process to ensure that the protection, contingency and response planning are balanced in relation to the threats, vulnerabilities, risk exposure and consequences of a potential cyber incident". (BIMCO et al., 2020)

In the paper of "Toward Automated Smart Ships: Designing Effective Cyber Risk Management" (2020) is mentioned that strong security in ships is necessary to consider both the security design at the time of ship building and the risk management once the ship is in service. Ship operators are also required to reduce security risks based on these guidelines. (Furumoto et al., 2020)

In the paper of "System-of-systems threat model" is defined criteria to assess the characteristics of the various threat models and their suitability for use. It is recommended in the paper that the models should be clustered into groups best suited for either strategic planning, engineering and acquisition, or operations. There is no one model, nor a cohesive suite of models, well suited for use at these three different levels of detail, (Bodeau & McCollum, 2018).

The threat probability tree model (Hummelholm, 2019) is used in this paper for the ship systems cyber security analysing. It has been used to develop the threat analysing **online** method and finally the idea is that it will improve elements of the remote pilotage situation awareness.

The paper of S4VF research program "ePilotage System of Systems' Cyber Threat Impact Evaluation" has been earlier in ICCWS2021 conference. This paper continues threat evaluation affords in the technical/tactical level and emphasizes the importation of all systems in the ship being elements of cyber security and ship's security situation awareness as a part of response process to smart maritime security. The research question is related to the enhancement of ship's cyber security risks procedure and how ship's cyber security situation awareness factors can be combined and how the information sharing to other stakeholders can be done on the maritime fairway domain.

The paper is organized as follows: the second chapter introduces a list of ships' systems and threats, the third chapter presents a method for organisations' cooperation, the fourth chapter discusses on vulnerabilities and possibilities for evaluation of the probability of a successful attack. Conclusions finish the paper in chapter five.

## **2. Ship systems and cyber security threats**

"Cybertechnologies have become essential to the operation and management of numerous systems critical to the safety and security of shipping and protection of the marine environment" stated the International Maritime Organization, IMO, in 2017. Lack of cyber security at a ship, technical products or networking infrastructure may result in a breach of ICT or OT systems and thus can affect to ship operation. Vulnerabilities in ICT or OT systems can lead to cyber risks which should be addressed. Vulnerable ship systems could include, but are not limited to: (IMO, 2017)

- Bridge systems;
- Cargo handling and management systems;

- Propulsion and machinery management and power control systems;
- Access control systems;
- Passenger servicing and management systems;
- Passenger facing public networks;
- Administrative and crew welfare systems; and
- Communication systems.

In “The Guidelines on Cyber Security Onboard Ships” paper (Annex 1) is listed ship`s systems in detailed way that may include vulnerable systems, equipment and technologies. It is a summary of potentially vulnerable systems and data onboard ships to assist companies with assessing their cyber risk exposure. (BIMCO et al., 2020)

In our paper ePilotage “System of Systems’ Cyber Threat Impact Evaluation” we have used ATT&CK (Strom et al. 2018. Alexander, Belisle and Steele, 2020) as a common behaviour-focused adversary model to assess ePilotage (Kovanen, Pöyhönen and Lehto, 2021). According to the paper the main interests for the adversary’s desired impacts are listed in ICT (Information and Communication Technology) and ICS (Industrial Control System). Those are the end point effects that manifest after successful attack paths are completed in the operations of the ships.

By combining IMO requirements, Mitre’s ATT&CK ICT and ICS Impacts (Mitre, 2019 and 2020) of successful attacks and the threat probability tree model we can be used for threat analyses and assessment method to develop ship cyber security and maintain situation awareness.

### **3. Description of situation awareness in operating environment**

The most significant challenges in the cybersecurity situation awareness of digital infrastructure relate to the detection of vulnerabilities and operational anomalies in complex technical systems. (Kokkonen, 2016)

According to the dissertation “Cyber security management and development as part of a critical infrastructure organization - System Thinking” (Pöyhönen, 2020), a critical infrastructure organization considers it important to provide a real-time status picture from its information systems and data resources and as well as industrial control systems (ICS or OT). In other words, there are needs for real-time situation awareness is divided into two challenges, namely the need for a snapshot of ICT reserves on the one hand and industrial automation on the other.

The analyzing usually happens in centralized monitoring rooms (Security Operations Center, SOC) based on the situation awareness. In the monitoring room, the information coming from different sensors are aggregated and a situation-specific analysis is formed. In this process and based on the analysis the needed operations are launched. The organization’s possibilities to utilize the information gotten from the international or national operational networks relate to the analysing ability. The personnel’s capability to interpret the available observations correctly has a significant meaning in composing situation-specific analyses. (Pöyhönen, 2020)

The Observe – Orient – Decide – Act (OODA) loop decision cycle depends completely upon tactical, operational, and strategic agility as stated by Boyd (1995): “Without OODA loops we can neither sense, hence observe, thereby collect a variety of information for the above processes nor decide as well as implement actions in accord with those processes. Without OODA loops embracing all the above and without the ability to get inside other OODA loops (or other environments), we will find it impossible to comprehend, shape, adapt to, and in turn be shaped by an unfolding, evolving reality that is uncertain, ever changing, unpredictable.”

Tikanmäki and Ruoslahti (2019) conclude that to build shared situation awareness organizations need information from the environment to notice the events surrounding them, and to understand their impact on their activities (Tikanmäki & Ruoslahti, 2019). The OODA loop is a framework that can provide structure to the collaboration aiming to better Cyber Situational Awareness (CSA) (Pöyhönen et al., 2020). In cybersecurity trust networks, the OODA loop provides organizations with a method of operational situation awareness and mutual decision-making, the principle of which is shown in Figure 1. The article “Cyber Situational Awareness in Critical Infrastructure” (Pöyhönen et al., 2020) discusses the OODA loop from the above perspective.

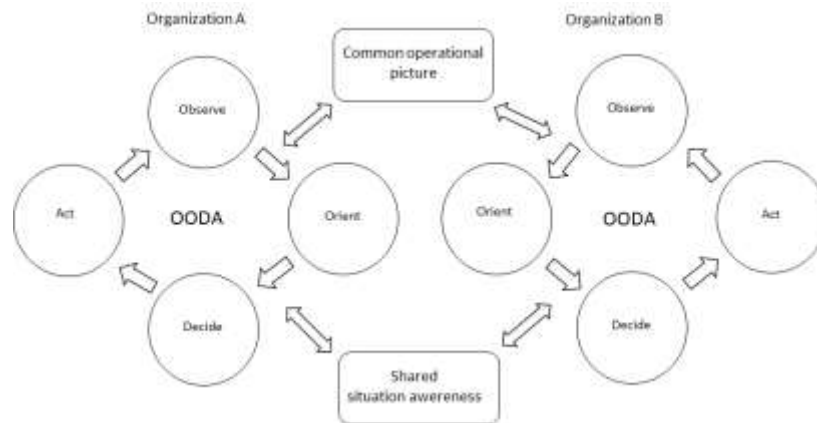


Figure 1: Decision making in complex environments. (Pöyhönen et al., 2020)

#### 4. Making and modelling threat analyses and situation awareness

When considering and modelling cyber threats and real-time situation picture of new autonomous vessels, a top-level description of autonomous vessel operating environments and their services is needed to perform analyses. Figure 2 shows the network architecture and responsibilities. In this context, there is talk of Cyber Secure Ship (SDB) and its systems. Making different communications zones inside of ships is important because vessels are connected to land-based communication systems and satellite systems. So, ship systems are also vulnerable to cyber-attacks in a similar way as our systems, applications and devices are. Attacks can come from either inside the ship or outside the ship. Hackers and cyber attackers can damage the operations of ships or stop ships from moving altogether.

Ships carry a lot of the materials we need on a daily basis and that our living environment needs. Ship transport must work well. Ship information systems are connected to each other as shown in Figure 3 via the ship's internal networks. Some of the ship's information systems are linked to port information systems, maritime authorities' information systems, the information systems of the various land-based stakeholders and the service providers' land-based services via fixed or mobile networks and when navigating at sea via satellites and some areas via mobile networks links, Figures 2 and 3. Figure 3 shows the different zones of the ship, which also means separating the information systems of the different areas of the ship with information security solutions and also separating the different spaces of the ship from each other because of security issues. One of the key factors in this separation of facilities and systems also comes from the need to physically separate systems because they are at different stages of technological development, as it is also possible to attack and deactivate ship systems through physical connections. There will be a lot of material on ships to be transported between ports, the transporting of goods is done by container ships as well as passenger ships. To all ship types, hackers and cyber attackers are able to bring different appropriate analysis tools to analyse the data networks inside the ship and retrieve vulnerabilities from the ship's information systems and the ship's propulsion and machinery management and power control systems.

With regard to the technologies used on ships and their management systems, the autonomous ships of the future (SDB), will be a complex entity using different communication solutions depending on whether the users and systems are connected to the ship's internal network or external networks. This operating environment will include information on the ship's machinery systems, navigation systems, radar systems, position and movement information of other ships and maritime surveillance systems, etc. From all of this information, we must form a situation picture from ship's systems and, in the case of information and communication systems, a situation picture of the cyber security situation in the ship.

In order to look more closely at and analyse the vulnerabilities of ship information and communication systems presented in the second chapter, we should divide ship systems into different zones (silos) based on their use, so that there is a clear picture of the whole. Figure 3 shows one principle, the architectural principle of upper-level autonomous ship systems, in which different ship systems and information and communication systems are divided into different zones.

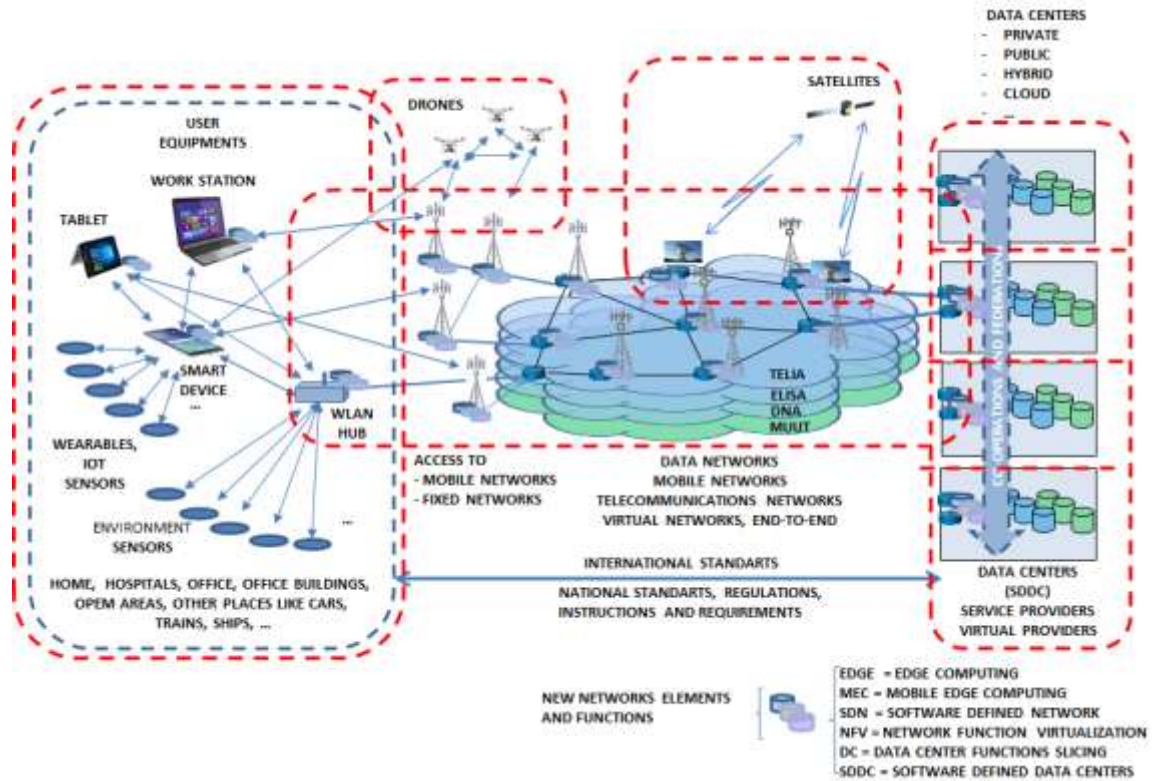


Figure 2: Network architecture (Hummelholm, 2018, p.526)

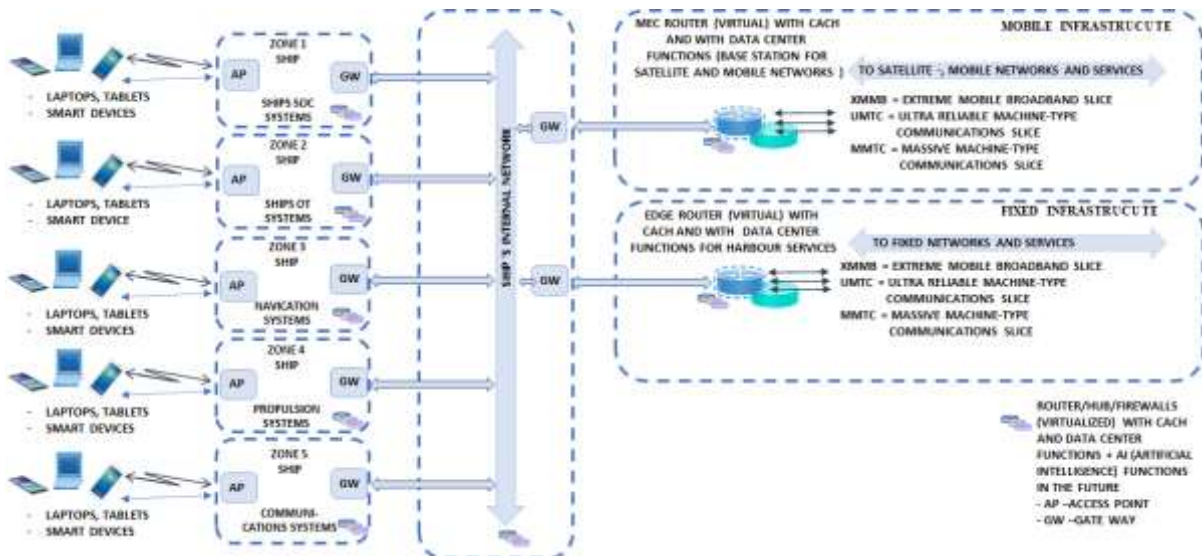


Figure 3: Autonomous ship's systems zones architecture

When we have an architectural view from users' smart devices and connections to different zones in a ship, we can take a closer look at the dependencies, risks, and cyber threats associated with the service chain we are used from our smart device when we are connected ship's internal bus. It is then possible to look at the dependencies, risks and cyber threats of each service chain and service up to the data center of the service provider (Figure 4). In cyber analyses, the Attack-tree model (Wang and Liu, 2014) and the Quality Function Deployment, QFD model for dependencies, (Chan and Wu, 2002) can be utilized in the reviews (Hummelholm, 2019, pp. 44-59).

From Figures 2-3 we can find out the service chains, which system or service is used at any given time and in which zone, to which service the smart devices used inside the ship is connected, which systems are currently controlled and what information is retrieved, etc. Probability calculations must be done for every sensors, every

smart devices, every AP and GW devices, GW devices to networks outside of ships, applications, software, systems, used protocols and data models in every zones.

As can be seen Figure 2-3, the access network environments contain a lot of wireless technology which is likely to increase in the future with we take in use mobile base stations (5G, 6G, 7G) and wireless local area networks to many places and also to inside of ships, which means small cellular networks inside of a ship. This increases the opportunities to connect to communications networks and access networks on board quickly and easily, but it also results in security and cyber security challenges in the use of on-board networks and the applications and services they contain. We must also take into account the risks and threats posed by these wireless networks when we carry out risk and threat analysis of ships. Once we have done the risk and threat analysis values, we can add the values obtained to Table 1. In Figure 4 we see attack tree model based on Figures 2 and 3 and that model is used here calculation of attack probabilities to different parts of ship information systems, command systems and steering systems. In Table 2 is presented the probability factors of risk (meaning of notations) in ships.

**Table 1:1** Threats and risks table (Hummelholm 2019, p.133)

Ref ID	Org	Functions	Category	Threat	Threat/Risk	Existing Control	Threat /risk level			Accept /reduce	Recom- mented control	Residual Threat/ Risk			CP
							L	C	R			L	C	R	
1 / AH	MC	Identify	Access	Foot- printing	Target Access	IDS /IPS	3	3	8	Reduce	EU- dir.	2	2	3	xx

Where L = Likelihood, C = Consequence, R = Risk, CP = Check Point

The Threats and Risks table is based on MITRE documents, Dragon1-open and JHS documents (Enterprise Architectures) and QFD model documents. The model is used to assist in the analysis of risks and threats in real networks and services and to present the probabilities of risks and threats also in audit and inspection situations and in organizations meetings given a situational picture of information and communication environment.

In Figure 4 the ship systems and sensors are divided to different sectors according to the technical structure. That are needed for the purpose to create necessary situation awareness from ship systems. The probability logic goes like:  $P(Ax_1)$  occurs if  $B_1$  or  $B_2$  or  $B_3... B_n$  occurs.  $P(Ax_2)$  occurs if  $C_1$  or  $C_2$  or  $C_3... C_n$  occurs.  $P(Dx_1)$  occurs if  $Ax_2$  or  $Ax_2$  or  $Ax_3$  occurs.  $P(Ex_1)$  occurs if  $P(Ax_1)$  occurs, or  $P(Ax_2)$  occurs. If we look at the sensors related to the steering of an autonomous vessel, we need to get the right information about all the sensors within the right limits so that the vessel can move. All sensors in this steering group must be on and on the right values. If anomalies occur, the vessel speed slow down or it stopped, and the information of anomalies is transferred automatically to the vessel’s SOC. From the device and zone level view of the probability examination can be extended to the system level as described below.

The National Institute of Standards and Technology (NIST) launched recommendations “Framework for Improving Critical Infrastructure Cybersecurity” (2018) for owners and operators of critical infrastructure to help them identify, assess, and manage cyber risks. The Framework Core part of guidance has a set of cybersecurity activities, outcomes, and informative references that are common across sectors and critical infrastructure. Elements of the Core provide detailed guidance will help an organization to align and prioritize its cybersecurity activities with its business/mission requirements, risk tolerances, and resources. The Framework Core consists of five concurrent and continuous Functions: Identify, Protect, Detect, Respond, Recover. There is mentioned that “The Framework is not a one-size-fits-all approach to managing cybersecurity risk for critical infrastructure”, thus we have used the framework to ship in part of S4F remote navigation research in this paper in the way described in Table 2. Probability to manage cyber security in ships can now be evaluated by using readiness to investigated elements of table. In this S4V research project the “Attack Identification” for ships is listed (ICT and ICS) (Kovanen et al., 2021) and “Protection” part is possible evaluate by using the status of “Protective Technology” readiness in ship use. In this paper for “Detection” purpose ship’s SOC (Security Operations Centre)

feature is recommended and from other research project CAN bus anomalies detection is also informed (Pöyhönen et al., 2019). “Countermeasure” activity count mainly on previous activities and in addition to that we suggest to concentrated to the “Communication and Analysis” in parts of ships cyber security countermeasures. In that sense real time Situation Awareness from ship and OODA-procedure are the key features with Conducting Response Planning and Mitigation procedures. All above element can be found in table 2 with other key activities according to NIST recommendations.

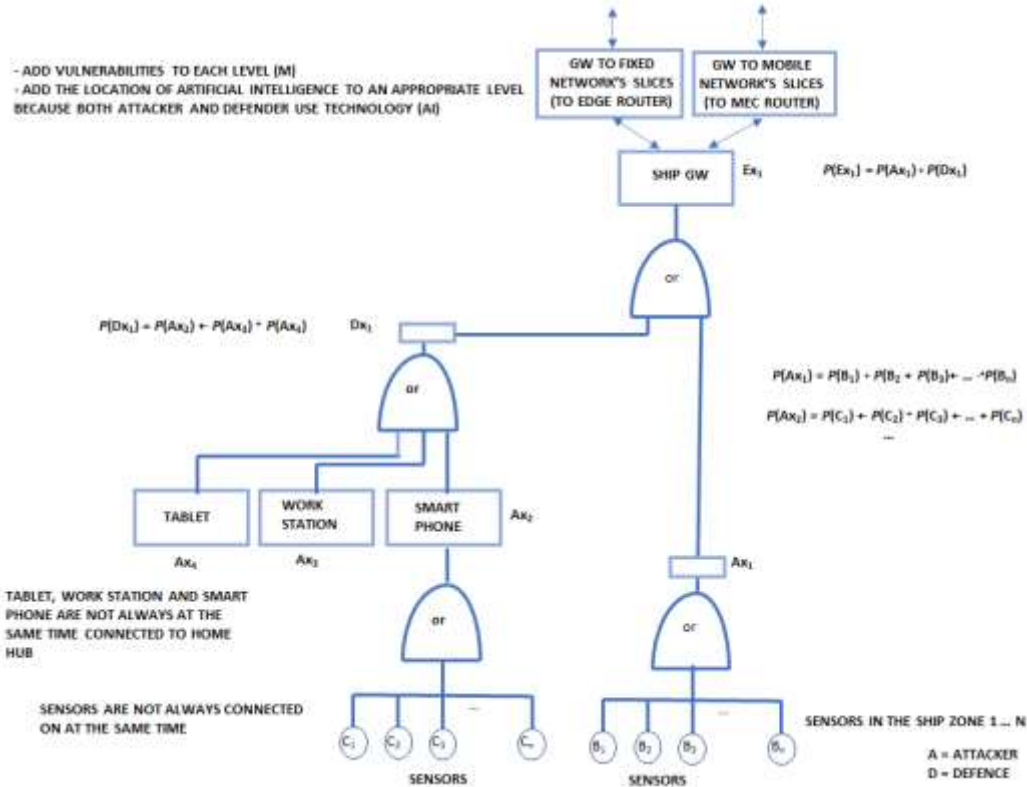


Figure 4: Attack tree model for ship’s systems to calculate cyber-attack probabilities (one zone)

Table 2: Meaning of Notations of ships risk probabilities (NIST, 2018; Kovanen et al., 2021; Pöyhönen et al., 2019; Pöyhönen et al., 2020)

ACTION	EXAMPLES	NOTATION
ATTACK IDENTIFICATION	ICT environment impacts ICS (OT) environment impacts	A
PROTECTION	Identity Management and Access Control, Awareness and Training, Data Security, Information Protection Processes and Procedures, Maintenance and Protective Technology (Port scan, FIREWALL, IDS, IPS, SIEM...)	P
DETECTION	Anomalies and Events, Security Continuous Monitoring and Detection Processes: a) ICT: SHIP SOC b) ICS (OT): CAN bus anomalies detecting based on message arrivals intervals	D
COUNTERMEASURE (RESPOND)	Conducting Response Planning, Communications and Analysis: a) Real time Situation Awareness from SHIP b) OODA-procedure Mitigation	M
RECOVERY	Recovery Planning, Improvements and Communications.	R



The probabilistic success of attacks ( $P(t)$ ) against an asset  $x$  in a ship can be calculated as follows adapting the paper of Threat Analysis of Cyber Attacks with Attack Tree+ (Wang and Liu, 2014)

$$P_{Ax}(t) = p_A(1 - P_P(t))(1 - p_D(t))(1 - p_M(t))(1 - p_R(t)) \quad (1)$$

where, as a function of time  $t$ , successful attack against a system  $x$ ,  $A_x$ , has attack success probabilities  $p_A$  reduced by defending mechanism: Protection  $P$ , Detection  $D$ , Countermeasure  $M$  and Recovery  $R$  having success probabilities  $p_P$ ,  $p_D$ ,  $p_M$  and  $p_R$  respectively.

A system can be faced with multiple types of attacks of which only one is required to succeed to continue the attack towards the gateway. In case of multiple attack types against an asset, the probability of an asset being successfully attacked is

$$P_{Ax}(t) = \sum_{i=1}^n p_{A_i}(t)((1 - (p_{P_i}(t))(1 - p_{D_i}(t))(1 - p_{M_i}(t))(1 - p_{R_i}(t))) \quad (2)$$

where, as a function of time  $t$ , successful attack  $A_x$  against a system  $x$ , has attack success probabilities  $p_A$  for different attacks  $A_i$ , reduced by defending mechanisms for the attack type  $i$  having success probabilities  $p_{P_i}$ ,  $p_{D_i}$ ,  $p_{M_i}$  and  $p_{R_i}$ .

An attack against the gateway can be direct or require one of the sub-assets to be compromised in order to launch an attack at the gateway. Direct attacks against the gateway can be expressed similarly as attacks against any other asset. Attacks requiring sub-asset compromise have a component of sub-asset success probabilities. The total probability for gateway compromise can be expressed by

$$P(t) = P_{AGW_{direct}}(t) + P_{AGW_x}(t) \sum_{i=1}^n P_{A_{x_i}}(t) \quad (3)$$

where  $P_{AGW_{direct}}$  is the successful attack probability of a direct gateway attack and  $P_{AGW_x}$  the probability of an attack requiring sub-asset compromise which's probability is expressed last.

## 5. Conclusions

The maritime sector is a vital part of the global economy, whether it is carrying cargo, passengers, or vehicles. Ships are becoming increasingly complex and dependent on the extensive use of digital and communications technologies throughout their operational life cycle. The maritime transportation system is a geographically and physically complex and diverse system consisting of waterways, ports, and intermodal landside connections. (DHS, 2014; Lehto, 2020)

Maritime transport is developing rapidly. Independent ships are evolving and coming into use in all seas. At the same time, the old generation of ships are moving at sea along with autonomous ships. The new autonomous vessels will include completely new information technology and new navigation systems, which must be ensured in all situations. Ships become very complex in their information systems, and they are also connected to various land systems. This creates a whole new type of system for maritime transport that needs ship's cyber security risks to be evaluate and managed in all situations. A probability evaluation to cyberattacks versus a probability to defend attacks is recommended approach to risk assessment. In that sense, it is very important to have real-time awareness of the cyber security situation about the systems, the ship's digital infrastructure, the navigation system, and the control system. In the case of ships, this claim is especially emphasized in the fairway navigation. The S4VF research program "ePilotage System of Systems' Cyber Threat Impact Evaluation" all those issues must be taken consideration.

There are needs for shared situation awareness between the remote ePilotage stakeholders and it can be done by using OODA-loop method. The events surrounding them are important to understand and all impacts to operations should be recognized. It is part of operational collaboration gaining better Cyber Situational Awareness (CSA) for every partner. The presented approach for modelling and evaluating risk probabilities produces information to be shared with OODA-loop method by controlled means between responsible stakeholders. This whole security process is highly recommended to be audit by the authorities.



## References

- Alexander, O., Belisle, M. and Steele, J. (2020). Mitre ATT&CK® for Industrial Control Systems: Design and Philosophy. BIMCO, CLIA, ICS, INTERCARGO, INTERMANAGER, INTERTANKO, IUMI, OCIMF and WORLD SHIPPING COUNCIL (2020). The Guidelines on Cyber Security Onboard Ships. Version 3. <https://www.ics-shipping.org/wp-content/uploads/2020/08/guidelines-on-cyber-security-onboard-ships-min.pdf>
- Bodeau, D. J. and McCollum, C. D., (2018). System-of-systems threat model. The Homeland Security Systems Engineering and Development Institute (HSSEDI) MITRE: Bedford, MA, USA.
- Boyd, J. R. (1995). The Essence of Winning and Losing. s.l.:s.n
- Chan, L. K. and Wu, M. L. (2002) Quality function deployment: a comprehensive review of its concepts and methods. Quality engineering, 15(1), 23-35.
- DHS. (2014). Sector Risk Snapshots, May 2014
- DIMECC Oy. (2020). Sea4Value / Fairway program (S4VF).
- Dragon1-open, EA Method / Visualization Standard, 'Enterprise Architecture Framework', <http://wiki.dragon1.org>.
- ENISA. (2019). PORT CYBERSECURITY. Good practices for cybersecurity in the maritime sector. November 2019.
- Furumoto, K., Kolehmainen, A., Silverajan, B., Takahashi, T., Inoue, D. and Nakao, K., (2020). Toward Automated Smart Ships: Designing Effective Cyber Risk Management. In 2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics) (pp. 100-105). IEEE.
- Hummelholm, A. (2018). "Cyber threat analysis in Smart City environments" 7 18<sup>th</sup> European Conference on Cyber Warfare and Security, 28 - 29 June 2019, University of Oslo, Norway, pages 523-532
- Hummelholm, A. (2019). Cyber Security and Energy Efficiency in the Infrastructures of Smart Societies Jyväskylä: University of Jyväskylä, 2019, 175 p.
- International Maritime Organization, IMO. (2017). Guidelines on Maritime Cyber Risk Management. MSC-FAL.1/Circ.3 5 July 2017
- JHS 179. Enterprise architecture planning, Modified date 2018-01-30, <http://www.jhs-suositukset.fi/web/guest/jhs/recommendations/179>.
- Kokkonen, T. (2016). Anomaly-based online intrusion detection system as a sensor for cyber security situational awareness system. Jyväskylä studies in computing, (251).
- Kovanen, T., Pöyhönen, J. and Lehto, M. (2021). ePilotage System of Systems' Cyber Threat Impact Evaluation. ICCWS2021.
- Lehto, M. (2020). Cyber Security in Aviation, Maritime and Automotive, in Pedro Diez, Pekka Neittaanmäki, Jacques Periaux, Tero Tuovinen, Jordi Pons-Prats (Edit.) Computation and Big Data for Transport, Springer 2020, pages 19-32. [https://doi.org/10.1007/978-3-030-37752-6\\_2](https://doi.org/10.1007/978-3-030-37752-6_2)
- Mitre. (2019). Impact. <https://attack.mitre.org/tactics/TA0040/>
- Mitre. (2020). Impact. <https://collaborate.mitre.org/attackics/index.php/Impact>
- National Institute of Standards and Technology, NIST. 2018. Framework for Improving Critical Infrastructure Cybersecurity, April 16, 2018
- Pöyhönen, J., Kotilainen, P., Kalmari J., Poikolainen, J., Neittaanmäki, P. 2019. Cyber security of vehicle CAN bus. ECCWS 2019: Proceedings of the 18th European Conference on Cyber Warfare and Security (pp. 354-363). Published by Academic Conferences and Publishing International Limited. Reading. UK
- Pöyhönen, J., Rajamäki, J., Ruoslahti, H. and Lehto, M. (2020). Cyber Situational Awareness in Critical Infrastructure Protection. Article approved 2nd March 2020 to Cyber Security of Critical Infrastructure 2020 (CYSEC2020) conference, October 27th, 2020 - October 28th, 2020. Dubrovnik. Croatia.
- Pöyhönen, J. (2020). Cyber security management and development as part of a critical infrastructure organization – System Thinking Jyväskylä: University of Jyväskylä, 2020, 236 p.
- QFD INSTITUTE, The official source for QFD, Quality Function Deployment (QFD), [http://www.qfdi.org/what\\_is\\_qfd/what\\_is\\_qfd.htm](http://www.qfdi.org/what_is_qfd/what_is_qfd.htm).
- Strom, B. E., Applebaum, A., Miller, D. P., Nickels, K. C., Pennington, A. G. and Thomas, C. B. (2018). *Mitre att&ck: Design and philosophy*. Technical report.
- Tikanmäki, I. and Ruoslahti, H. (2019). How Are Situation Picture, Situation Awareness, and Situation Understanding Discussed in Recent Scholarly Literature? In: A. S. a. J. F. Jorge Bernardino, ed. Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. Portugal: SCITEPRESS – Science and Technology Publications, Lda., pp. 419-426.
- Wang, P. and Liu, J. C. (2014). Threat analysis of cyber-attacks with attack tree+. Journal of Information Hiding and Multimedia Signal Processing, 5(4).

# A Review of National Cyber Security Strategies (NCSS) Using the ENISA Evaluation Framework

Angela Jackson-Summers

U.S. Coast Guard Academy, New London, USA

[angela.g.jackson-summers@uscga.edu](mailto:angela.g.jackson-summers@uscga.edu)

DOI: 10.34190/EWS.21.040

**Abstract:** During the COVID-19 pandemic, cyber threats have continued to increase prompting increased awareness and preparedness from malicious security attacks. In April 2020, a joint alert, *COVID-19 Exploited by Malicious Cyber Actors*, was issued from the United Kingdom's National Cyber Security Centre (NCSC) and the United States Department of Homeland Security (DHS) Cybersecurity and Infrastructure Security Agency (CISA). Given the advancement of cyber threats, especially during our current COVID-19 pandemic, national cyber security strategy (NCSS) effectiveness remains important. The purpose of this study is intended to determine existing challenges in NCSS effectiveness. Using textual analysis, the national cyber security strategies (NCSS') for 18 countries were evaluated against the European Union Agency for Cybersecurity's (ENISA) Evaluation Framework dated November 2014. The results of this study reflected a need for maturation and timelier updates to NCSS'. Gaps and recommendations are discussed and provide future research considerations. This study's contribution provides additional insights to researchers, practitioners, and other stakeholders focused on national cyber security policy, related strategies, and risk management practices.

**Keywords:** cyber security, national strategy, national security, security policy, risk management

---

## 1. Introduction

With increasing cyber security threats in 2020, a joint alert was issued by the United Kingdom's National Cyber Security Centre (NCSC) and the United States Department of Homeland Security (DHS) Cybersecurity and Infrastructure Security Agency (CISA) (Cybersecurity & Infrastructure Security Agency, 2020). The readiness of some nations to address cyber security threats are reliant upon their most current national cyber security strategy (NCSS), including their execution in use of security resources and related security capabilities. So, how ready are nations now to address these advancing cyber security threats?

Early literature has shown existing NCSS differences, such approaches and focal cyber security threat aspects, resulting in reported weaknesses (Luijff, Besseling, & de Graaf, 2013). Today, the European Union Agency for Network and Information (ENISA), <http://www.enisa.europa.eu/>, provides a NCSS evaluation framework resulting from a study designed to help address cyber security specific objectives as described below (Liveri & Sarri, 2014).

- Cyber defence policies/capabilities development
- Cyber resilience achievements
- Cybercrime reduction
- Industry cyber security support
- Critical information infrastructures' security

While dated over five years ago, the basis of ENISA's NCSS evaluation framework was empirically driven (Liveri & Sarri, 2014). With the above defined NCSS cyber security specific objectives, three intended purposes (stated below) of NCSS were also shared (Liveri & Sarri, 2014). Those three specific purposes helped better frame the NCSS as strategically serving documents that specifically support the national and international policy landscape (Liveri & Sarri, 2014).

- Holistic governmental alignment
- Focal and structural support for stakeholder dialogue
- Partnership priorities among Member States and international communities

Given the increasing cyber threats nations face today, the purpose of this study is to examine existing NCSS for their effectiveness. Using ENISA's Evaluation Framework, the research question addressed by this study is as follows.

- What challenges exist toward NCSS effectiveness?

The eighteen (18) countries, Australia, Canada, Czech Republic, Estonia, France, Germany, India, Japan, Lithuania, Luxembourg, Romania, The Netherlands, New Zealand, South Africa, Spain, Uganda, the United Kingdom, and the United States addressed in Luijff, Besseling, & Graaf's (2013) study serve as this study's case. Some of those countries also served as part of the sample used in the study that rendered ENISA's November 2014 NCSS Evaluation Framework. The sample was selected by using the same countries identified in the Luijff, Besseling, & Graaf's (2013) study, which provided baseline perspectives of their cyber security objectives and maturity levels prior to this study being performed. The most recent national cyber security strategy (NCSS) in place as of March 31, 2020 was captured and examined in this study for the selected sample. See Table 1 below for a listing of NCSS' by country. Textual analysis of keywords was used to support the completion of this study's examination.

**Table 1:** Sample National Cyber Security Strategies (NCSS) by country

Country	Year Published	NCSS Links
Australia	2016	<a href="https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/Australia_2016_Cyber-Strategy.pdf">https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/Australia_2016_Cyber-Strategy.pdf</a>
Canada	2018	<a href="https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ntnl-cbr-scrtr-strtg/index-en.aspx">https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ntnl-cbr-scrtr-strtg/index-en.aspx</a>
Czech Republic	2015	<a href="https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/National%20Cyber%20Security%20Strategy%20-%20Czech%20Republic.pdf">https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/National%20Cyber%20Security%20Strategy%20-%20Czech%20Republic.pdf</a>
Estonia	2014	<a href="https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/Estonia_Cyber_security_Strategy.pdf">https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/Estonia_Cyber_security_Strategy.pdf</a>
France	2018	<a href="http://www.sgdsn.gouv.fr/uploads/2018/02/20180206-np-revue-cyber-public-v3.3-publication.pdf?bcsi_scan_858c91d0398e8bd7=0&amp;bcsi_scan_filename=20180206-np-revue-cyber-public-v3.3-publication.pdf">http://www.sgdsn.gouv.fr/uploads/2018/02/20180206-np-revue-cyber-public-v3.3-publication.pdf?bcsi_scan_858c91d0398e8bd7=0&amp;bcsi_scan_filename=20180206-np-revue-cyber-public-v3.3-publication.pdf</a>
Germany	2016	<a href="https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map/strategies/cyber-security-strategy-for-germany/@@download_version/5f3c65fe954c4d33ad6a9242cd5bb448/file_en">https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map/strategies/cyber-security-strategy-for-germany/@@download_version/5f3c65fe954c4d33ad6a9242cd5bb448/file_en</a>
India	2013	<a href="https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/India_2013_National_cyber_security_policy-2013%281%29.pdf">https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/India_2013_National_cyber_security_policy-2013%281%29.pdf</a>
Japan	2018	<a href="https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/cs-senryaku2018-en.pdf">https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/cs-senryaku2018-en.pdf</a>
Lithuania	2018	<a href="https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/LRV+818+National+Cyber+Scurity+Strategy+%28Lithuania%29.pdf">https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/LRV+818+National+Cyber+Scurity+Strategy+%28Lithuania%29.pdf</a>
Luxembourg	2018	<a href="https://hcn.gouvernement.lu/dam-assets/fr/publications/brochure-livre/national-cybersecurity-strategy-3/national-cybersecurity-strategy-iii-en.pdf">https://hcn.gouvernement.lu/dam-assets/fr/publications/brochure-livre/national-cybersecurity-strategy-3/national-cybersecurity-strategy-iii-en.pdf</a>
Romania	2013	<a href="https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map/strategies/cyber-security-strategy-in-romania/@@download_version/1b41c7f470b14b52be67866e84007f87/file_en">https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map/strategies/cyber-security-strategy-in-romania/@@download_version/1b41c7f470b14b52be67866e84007f87/file_en</a>
The Netherlands	2018	<a href="https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map/strategies/national-cyber-security-strategy-1/@@download_version/82b3c1a34de449f48cef8534b513caea/file_en">https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map/strategies/national-cyber-security-strategy-1/@@download_version/82b3c1a34de449f48cef8534b513caea/file_en</a>
New Zealand	2019	<a href="https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/nz-cyber-security-strategy-december-2015.pdf">https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/nz-cyber-security-strategy-december-2015.pdf</a>
South Africa	2010	<a href="https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/SouthAfrica_2010_100219cybersecurity.pdf">https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/SouthAfrica_2010_100219cybersecurity.pdf</a>
Spain	2019	<a href="https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map/strategies/the-national-security-strategy/@@download_version/5288044fda714a58b5ca6472a4fd1b28/file_en">https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map/strategies/the-national-security-strategy/@@download_version/5288044fda714a58b5ca6472a4fd1b28/file_en</a>
Uganda	2014	<a href="https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/Uganda_2014_National%20Information%20Security%20Policy%20v1.0_0.pdf">https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/Uganda_2014_National%20Information%20Security%20Policy%20v1.0_0.pdf</a>

Country	Year Published	NCSS Links
United Kingdom	2016	<a href="https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/national_cyber_security_strategy.pdf">https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/national_cyber_security_strategy.pdf</a>
United States	2018	<a href="https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf">https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf</a>

Text analysis is heavily rooted in the social sciences producing pattern outcomes that offer an alternative approach to seek embedded meaning (Benard, & Ryan, 1998) in artifacts, such as documents and other artifacts. Text analysis has provided further value in insightful, alternative ways to handle text, linguistically, as data (Benard, & Ryan, 1998). For example, in prior studies, text analysis has helped classification efforts (Cheong, Yoon, Cho, & No, 2020) and support topic modeling in policy studies (Isoaho, Gritsenko, & Mäkelä, 2021) to share differing perspectives and meaning.

In the next section, this study’s research approach to examining NCSS for their effectiveness is described. In the following section, the major findings are provided from the study’s examination performed. Finally, a discussion of the study’s major findings, including gaps and recommendations, and implications for research and practice serves as the conclusion of this paper.

## 2. Methods

Given the guiding principles of ENISA’s NCSS Evaluation Framework for existing and newly developed cyber security strategies, the framework served as the foundational model in this study in which to evaluate current national cyber security strategies (NCSS). Not all countries selected for this study’s sample are members of the European Union (EU). The NCSS Evaluation Framework presented a model in which to establish a baseline in identifying challenges toward NCSS effectiveness.

This study’s sample is based on the NCSS’ of eighteen (18) countries that were examined in a 2013 study (Luijff, Besseling, & de Graaf, 2013). The 18 countries included Australia, Canada, Czech Republic, Estonia, France, Germany, India, Japan, Lithuania, Luxembourg, Romania, The Netherlands, New Zealand, South Africa, Spain, Uganda, the United Kingdom, and the United States. The NCSS sample has previously been reported as having differences among their cyber security objectives, approaches, definition use of key terms, and readiness maturity (Luijff, Besseling, & de Graaf, 2013).

Prior literature has shown that textual analysis can offer insights and help in classification efforts, when needed (Cheong, et al., 2020; Isoaho, et al., 2021). Performing textual analysis in this study was intended to help categorize strategic actions that may be focal to a country’s cyber security specific objectives. See Table 2 below for those keywords identified and used for textual analysis.

**Table 2:** Keywords Identified by NCSS Evaluation framework’s cyber security specific objectives (Liveri & Sarri, 2014) and used in textual analysis performed

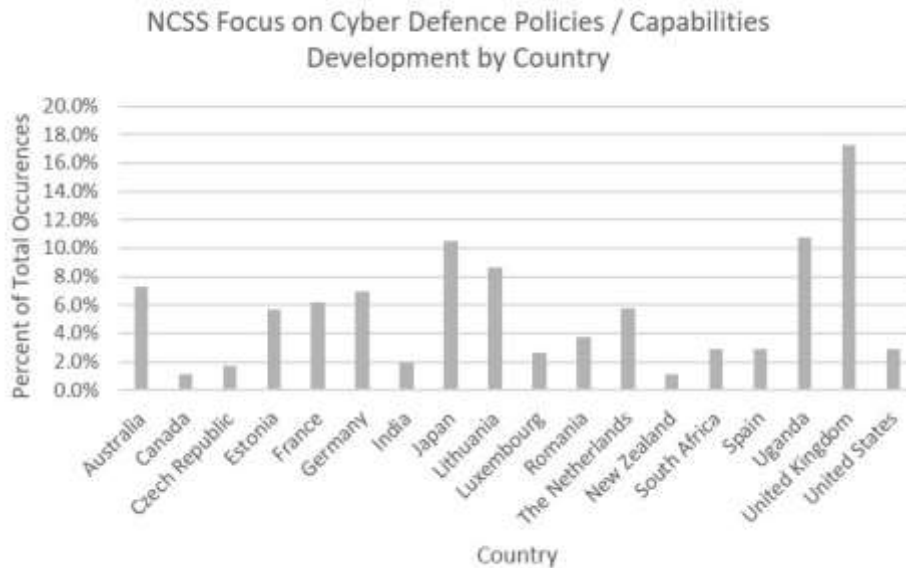
NCSS Evaluation Framework’s Cyber Security Specific Objectives (Liveri & Sarri, 2014)	Keywords
Cyber defence policies/capabilities development	Defence, defense, policy, policies, capable, capability, capabilities
Cyber resilience achievements	Resilient, resilience
Cybercrime reduction	Cybercrime, crime
Industry cyber security support	Security, support
Critical information infrastructures security	Critical, data, information, infrastructure, infrastructures

## 3. Results

For this study, a sample of eighteen (18) countries, Australia, Canada, Czech Republic, Estonia, France, Germany, India, Japan, Lithuania, Luxembourg, Romania, The Netherlands, New Zealand, South Africa, Spain, Uganda, the United Kingdom, and the United States were selected. The sample was selected by using the same countries identified in the Luijff, Besseling, & Graaf’s (2013) study, and capturing their most recent national cyber security strategy (NCSS) in place as of March 31, 2020. Using NVivo 12 Plus, the following results were categorized by the NCSS Evaluation Framework’s cyber security specific objectives (Liveri & Sarri, 2014).

### 3.1 Cyber defence policies / capabilities development

The first focal cyber security specific objective relates to cyber defence policies and the development of capabilities. Three (3) countries, Japan (10.5%), Uganda (10.8%), and the United Kingdom (17.3%), representing 16.7%, of the 18 countries showed greater emphasis on their efforts to address cyber defence policies and development their cyber readiness capabilities. However, four (4) countries, Canada (1.1%), Czech Republic (1.7%), India (2.0%), and New Zealand (1.1%), representing 22.2%, of the 18 countries showed minimal interest in reflecting their efforts to drive their national cyber defence policies and to build related cyber defence capabilities. Figure 1 below depicts the 18 countries' NCSS focus on cyber defence policies and capabilities development.



**Figure 1:** National Cyber Security Strategies (NCSS) focal cyber security specific objective 1: Cyber defence policies / capabilities development by country

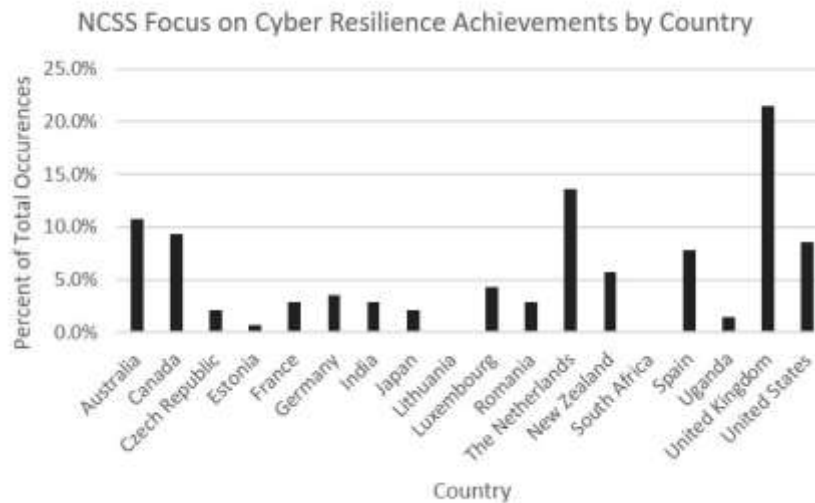
For more detail, see Table 3 below.

**Table 3:** Keyword results for NCSS evaluation framework's cyber security specific objective (cyber defence policies / capabilities development) by country

Country	Keywords Total Count Occurrences							Total
	Capable	Capability	Capabilities	Defence	Defense	Policy	Policies	
Australia		17	19	16		17	2	71
Canada		1	6	1		3		11
Czech Republic		1	4	10		1	1	17
Estonia	1	3	14	29		7	1	55
France	2	2	7	33	13	3		60
Germany		1	14	26		27		68
India	2		2	1		9	5	19
Japan	6	7	29		22	30	8	102
Lithuania	2		20	53		9		84
Luxembourg		1		15		2	8	26
Romania			14	1	10	3	8	36
The Netherlands	6	1	28	8		13		56
New Zealand	4		2			4	1	11
South Africa			1			25	2	28
Spain	1	1	12	7		5	2	28
Uganda				7		74	24	105
United Kingdom	2	46	57	38		18	7	168
United States		1	11		4	7	5	28
Total	26	82	240	245	49	257	74	973

### 3.2 Cyber resilience achievements

The second focal cyber security specific objective relates to cyber resilience achievements. Three (3) countries, Australia (10.7%), the Netherlands (13.6%), and the United Kingdom (21.4%), representing 16.7%, of the 18 countries showed greater emphasis on their cyber resilience achievements. However, four (4) countries, Estonia (0.7%), Lithuania (0.0%), South Africa (0.0%), and Uganda (1.4%), representing 22.2%, of the 18 countries showed minimal to no interest focusing on their cyber resilience achievements. Figure 2 below depicts the 18 countries' NCSS focus on cyber resilience achievements.



**Figure 2:** National Cyber Security Strategies (NCSS) focal cyber security specific objective 2: Cyber resilience achievements by country

For more detail, see Table 4 below.

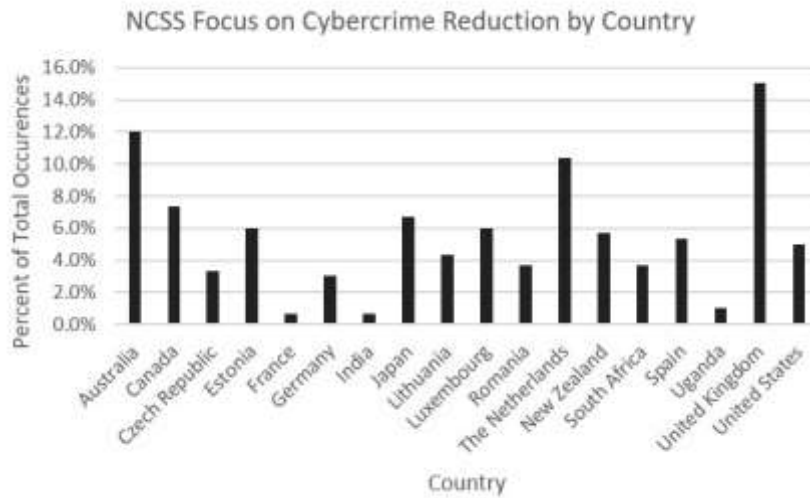
**Table 4:** Keyword results for NCSS evaluation framework's cyber security specific objective (cyber resilience achievements) by country

Country	Keywords		Total
	Resilient	Resilience	
Australia	8	7	15
Canada	3	10	13
Czech Republic	2	1	3
Estonia		1	1
France		4	4
Germany	1	4	5
India	1	3	4
Japan		3	3
Lithuania			0
Luxembourg	1	5	6
Romania		4	4
The Netherlands	9	10	19
New Zealand	5	3	8
South Africa			0
Spain		11	11
Uganda		2	2
United Kingdom	15	15	30
United States	3	9	12
Total	48	92	140

### 3.3 Cybercrime reduction

The third focal cyber security specific objective shares emphasis on cybercrime reduction. Three (3) countries, Australia (12.0%), the Netherlands (10.4%), and the United Kingdom (15.1%), representing 16.7%, of the 18

countries showed greater emphasis on cybercrime reduction. However, four (4) countries, France (0.7%), India (0.7%), and Uganda (1.0%), also representing 16.7%, of the 18 countries showed minimal interest on cybercrime reduction. Figure 3 below depicts the 18 countries' NCSS focus on cybercrime reduction.



**Figure 3:** National Cyber Security Strategies (NCSS) focal cyber security specific objective 3: Cybercrime reduction by country

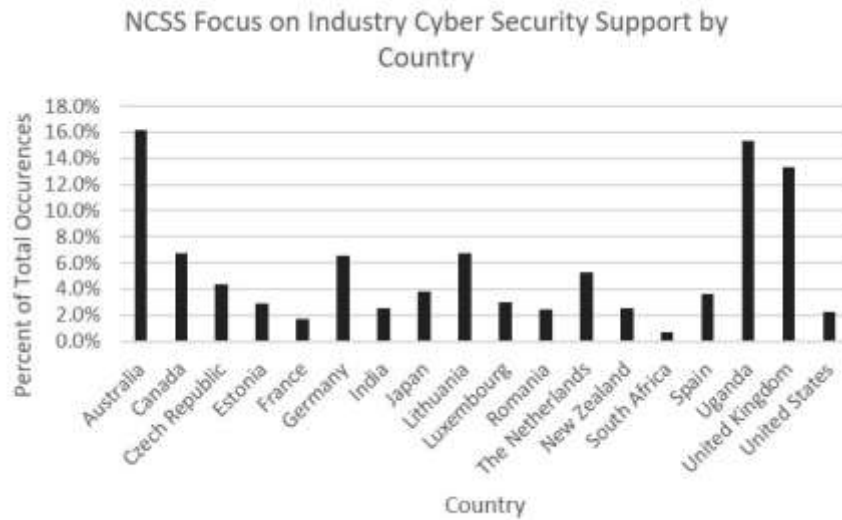
For more detail, see Table 5 below.

**Table 5:** Keyword results for NCSS evaluation framework's cyber security specific objective (cybercrime reduction) by country

Country	Keywords		Total
	Crime	Cybercrime	
Australia	5	31	36
Canada	3	19	22
Czech Republic	3	7	10
Estonia		18	18
France		2	2
Germany	8	1	9
India	2		2
Japan	5	15	20
Lithuania	4	9	13
Luxembourg	1	17	18
Romania	2	9	11
The Netherlands	10	21	31
New Zealand	4	13	17
South Africa		11	11
Spain	4	12	16
Uganda	3		3
United Kingdom	43	2	45
United States	5	10	15
Total	102	197	299

### 3.4 Industry cyber security support

The fourth focal cyber security specific objective addresses the importance of industry cyber security support. Three (3) countries, Australia (16.2%), Uganda (15.3%), and the United Kingdom (13.3%), representing 16.7%, of the 18 countries showed greater emphasis on industry cyber security support. However, two (2) countries, France (1.7%), and South Africa (0.7%), also representing 11.1%, of the 18 countries showed less emphasis on industry cyber security support. Figure 4 below depicts the 18 countries' NCSS focus on industry cyber security support.



**Figure 4:** National Cyber Security Strategies (NCSS) focal cyber security specific objective 4: Industry cyber security support by country

For more detail, see Table 6 below.

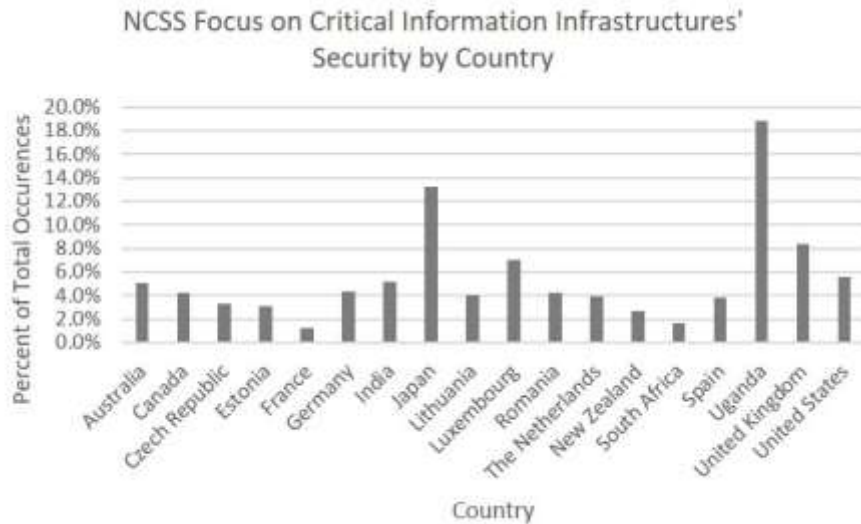
**Table 6:** Keyword results for NCSS evaluation framework’s cyber security specific objective (industry cyber security support) by country

Country	Keywords		Total
	Security	Support	
Australia	496	31	527
Canada	208	11	219
Czech Republic	133	9	142
Estonia	84	9	93
France	41	14	55
Germany	203	11	214
India	81	2	83
Japan	110	13	123
Lithuania	217	3	220
Luxembourg	94	4	98
Romania	76	3	79
The Netherlands	167	3	170
New Zealand	80	3	83
South Africa	17	5	22
Spain	106	11	117
Uganda	489	9	498
United Kingdom	394	39	433
United States	61	13	74
Total	3,057	193	3,250

### 3.5 Critical information infrastructures’ security

The last focal cyber security specific objective addresses critical information infrastructures’ security. Two (2) countries, Japan (13.3%), and Uganda (18.8%), representing 11.1%, of the 18 countries showed greater emphasis on the security of their national critical information infrastructures. However, two (2) countries, France (1.2%), and South Africa (1.6%), also representing 11.1%, of the 18 countries showed less emphasis on the security of their national critical information infrastructures. Figure 5 below depicts the 18 countries’ NCSS focus on critical information infrastructures’ security.





**Figure 5:** National Cyber Security Strategies (NCSS) focal cyber security specific objective 5: Critical information infrastructures' security by country

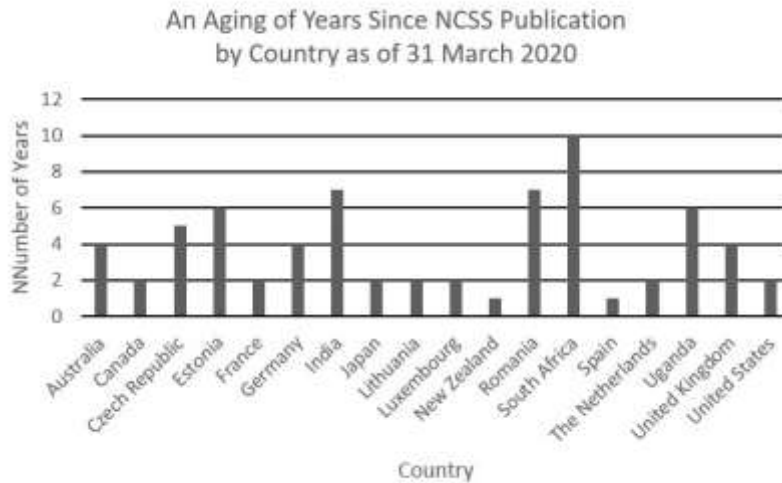
For more detail, see Table 7 below.

**Table 7:** Keyword results for NCSS evaluation framework's cyber security specific objective (critical information infrastructures' security) by country

Country	Keywords					Total
	Critical	Data	Information	Infrastructure	Infrastructures	
Australia	21	17	62	9		109
Canada	16	14	44	18		92
Czech Republic	4	18	41	9		72
Estonia	11	7	31	16	1	66
France	4	7	12		3	26
Germany	7	23	47	9	8	94
India	17	5	61	29		112
Japan	36	48	161	39	1	285
Lithuania	8	24	46	9		87
Luxembourg	10	15	91	21	14	151
Romania	9	4	32	32	14	91
The Netherlands	14	22	33	16		85
New Zealand	12	9	25	8	3	57
South Africa	5	5	20	5		35
Spain	11	11	42	6	13	83
Uganda	58	54	234	55	3	404
United Kingdom	20	58	71	30	1	180
United States	24	15	36	45	1	121
Total	287	356	1,089	356	62	2,150

### 3.6 Summary

Of the 18 NCSS' examined, the results varied. One significant result was reflected in the aging of the last NCSS by country published as of 31 March 2020 that shows that nine (9), 50%, of the 18 NCSS' were published within the past 3 years. Eight (8), 44.4%, of the 18 NCSS' were published 3-7 years ago, and one (1), 5.6%, NCSS was published over 8 years ago. For a depiction of the aging results by country, see Figure 6 below.



**Figure 6:** An aging of years since National Cyber Security Strategies (NCSS) publication by country as of 31 March 2020

#### 4. Discussion

While the above reported findings reflect varying gaps and perceived challenges that countries face in showcasing their NCSS efforts to drive the five cyber security specific objectives, a major gap appears to be the lack of maturation and timely NCSS efforts. Some countries' NCSS efforts do reflect such efforts as being commensurate with increasing cyber threats, and technological advancements, including their relative complexities. However, if ENISA's NCSS Evaluation Framework was designed to help address the identified cyber security specific objectives, having greater evidence of a country's focal efforts pertaining to these cyber security specific objectives is recommended.

Upon further review of the sample NCSS', another major gap noted was the inability to acquire personnel with the necessary skills to support a country's efforts to address their reported cyber security challenges. Some of those reported challenges included the increased number of online devices attributed to the Internet of Things (IoT). Other challenges also noted legal and law enforcement inability to protect against malicious actions, and the continued growth of breaches, cybercrimes, and disruption of essential services in varying areas within these countries.

One of the major limitations of this study is the limited number of sample country NCSS' examined. There are future opportunities to provide additional research on the gaps and challenges that countries face in maturing their NCSS efforts and publishing timelier NCSS'. This study's contribution mainly offers researchers and practitioners an alternative perspective in how to identify NCSS' needs of maturity and timely updates as well as a greater focus on cyber security specific objectives.

#### Acknowledgements

Research funding was provided by the Center for Advanced Studies (CAS) at the United States Coast Guard Academy in New London, CT USA.

#### References

- Bernard, H. R., & Ryan, G., 1998. Text analysis. *Handbook of methods in cultural anthropology*, 595-645.
- Cheong, A., Yoon, K., Cho, S. & No, W. G., 2020. Classifying the contents of cybersecurity risk disclosure through textual analysis and factor analysis. *Journal of Information Systems*, doi: <https://doi.org/10.2308/ISYS-2020-031>
- Cybersecurity & Infrastructure Security Agency, 2020. *CISA.gov National Cyber Awareness System Alert (AA20-099A): COVID-19 exploited by malicious cyber actors*. [Online] Available at: <https://us-cert.cisa.gov/ncas/alerts/aa20-099a>
- ENISA, 2012. *National Cyber Security Strategies: Practical Guide on Development and Execution*, s.l.: ENISA.
- Isoaho, K., Gritsenko, D., & Mäkelä, E., 2021. Topic modeling and text analysis fo qualitative policy research. *Policy Studies Journal*, 49(1), 300-324.
- Liveri, D. & Sarri, A., 2014. *An evaluation framework for national cyber security strategies*, s.l.: Heraklion: ENISA.
- Luijff, E., Besseling, K. & de Graaf, P., 2013. "Nineteen national cyber security strategies". *International Journal of Critical Infrastructure Protection*, 9(1/2), pp. 3-31.

# Some Cybersecurity Governance Imperatives in Securing the Fourth Industrial Revolution

Victor Jaquire<sup>1</sup>, Petrus Duvenage<sup>1</sup> and Sebastian von Solms<sup>2</sup>

<sup>1</sup>Academy of Computer Science and Software Engineering, University of Johannesburg, South Africa

<sup>2</sup>Centre for Cyber Security, University of Johannesburg, South Africa

[victorJ@uj.ac.za](mailto:victorJ@uj.ac.za)

[pcduvenage@uj.ac.za](mailto:pcduvenage@uj.ac.za)

[basievs@uj.ac.za](mailto:basievs@uj.ac.za)

DOI: 10.34190/EWS.21.056

**Abstract:** The fourth industrial revolution (4IR) will “transform the workplace from the existing pattern to the “human centred” characteristics” and just as there is a “tendency to merge man and machine to shorten the distance between natural sciences, humanities and social sciences, the same is expected to happen with science and technology” (Scepanović 2019). Scepanović (2019) further indicates that these processes will require a “shift to interdisciplinary teaching, research and innovation, education and in particular higher education”, and that “the 4IR has a special role, being a complex, dialectical and exciting activity which promises to transform society for the better”. In contrast, Onik, Kim and Yang (2019) state that the 4IR will “expose the maximum personal information the world has ever seen” and observe that “although people are considered as an asset, several recent information leaking incidents have shaken the whole world in perspective of data privacy”. Also within the 4IR context, Sander (2019) refers to the realities of cyber-attacks, indicating that “Peacetime cyber-attacks are destructive cyber operations, encompassing acts undertaken by a State – or actors whose conduct is attributable to a State under international law – that uses cyber capabilities to alter, disrupt, degrade or destroy the computer systems or networks of a foreign State, or the information or programs resident in those systems or networks” It therefore becomes clear that there should be cyber security considerations with regard to the envisaged outcomes and realities culminating from the 4IR. This paper contributes to a series of previous papers on cyber security and cyber counterintelligence (CCI). Its primary aim is to add to the burgeoning academic discourse on emerging cyber concerns through a discussion of some cyber security governance imperatives in relation to the 4IR. To this end, the paper highlights some positive and negative realities (outcomes) of the 4IR, including some background to the 4IR’s impact. We then proceed with examining some potentially dire consequences flowing from the 4IR. Finally, the paper advances a proposition on addressing these consequences. Our proposition is specifically focussed on three imperatives to an effective cybersecurity governance approach in securing the 4IR, namely: ‘strategy’, ‘intelligence and counterintelligence’ as well as ‘capacity building and skills development’.

**Keywords:** cyber security, cyber counterintelligence, governance, cyber threat intelligence, defensive and offensive cybersecurity, fourth industrial revolution

---

## 1. Introduction

Since before the notion of the 4<sup>th</sup> industrial revolution (4IR) fully gained traction a stream of opinions, guidelines and other contributions emerged in the academic realm, in discussion of this theme. The idea of this latest revolution brings with it mixed sentiments and realities, especially within some developing (and developed) countries still struggling to fully get to grips with the third industrial revolution.

Global events of late, assisted to hasten the implementation of 4IR concepts, fuelling the rapid cyberfication of systems and services to enable governments, companies and individuals to exist, transact and thrive online, owing to fears of COVID 19 contamination within the physical world (de Castro Sobrosa Neto *et al* 2020).

Not only does this precipitous uptake of the virtual sphere bring with it a renewed emphasis on the realities that a virtual lifestyle generates, but the magnitude and abrupt acceptance of this new-normal brings with it sets of outcomes ranging from the ‘almost very good’ to the ‘almost very bad’. The 4IR concurrently holds consequences potentially so dire (‘almost very ugly’) that it is imperative to address these through effective cybersecurity governance. In addressing this problem statement, the rest of paper is structured as follows:

- *Section 2: Some Things Almost Very Good* briefly outlines some positive outcomes of the 4IR.
- *Section 3: Some Things Almost Very Bad* examines some potentially negative aspects of the 4IR such as the interrelated changes in technology adoption, privacy and job roles.

- Section 4: Some Almost Very Ugly Consequences highlights some potentially dire consequences resulting from the 4IR.
- Section 5: Some Cybersecurity Governance Imperatives and the 4IR expounds the following three elements we identified as critical to effectively mitigate adverse consequences namely:
  - *Strategy,*
  - *Intelligence and Counterintelligence, and*
  - *Capacity Building and Skills Development.*

## **2. Some things, almost very good**

The assigning of descriptors such as ‘very good’, ‘very bad’ and even more so ‘very ugly’ is admittedly subjective and determined by a particular reference framework. Nonetheless, and risking oversimplification, aspects such as opportunities, growth, innovation and education would generally be deemed as ‘very good’, whereas the aspects denoting of cyber insecurity could be seen as ‘very bad’.

From a business perspective, Scepanović (2018) posits that “The Fourth Industrial Revolution (4IR) forces humans to encourage creative thinking about the manufacturing processes, value chain, and customer service processes”. Penker and Khoh (2018) state that “At the dawn of the fourth industrial revolution, innovation is key for organisations that aspire to grow and retain their business relevance”. Anser *et al* (2020) and Bigerna *et al* (2021) resonate this by asserting that “The 4th industrial revolution (4IR) is expected to promote a radical socioeconomic transformation” but further note that “Future generations will have to grasp the potential of the 4IR to reduce and, hopefully, eliminate the environmental, social, and economic damages of previous industrial revolutions”.

It seems that the digital age has certainly provided more options to the small, medium and large enterprises when it comes to doing business, managing certain risks and providing a more level playing field, in at least for as much as the access to certain online technologies are concerned. Muriel-Pera *et al* (2018) state that “democratization of technology is positive” and go further in indicating that “Companies, regardless of their size can access technologies without requiring big infrastructures or specialized knowledge to manage them out”. Increased productivity, efficiency in processes, enhanced decision-making with data tools, as well as the acceleration of research and product development are some fruits of such technological democratisation.

For the individual, there seems to be, as an example, more opportunities to work online from home. This is not a new concept, but definitely something that is, and could become more of the norm as the 4IR and the idea of a digital lifestyle takes more prevalence. This provides opportunities to individuals, who otherwise would not have been able to travel (for whatever reason) to either enter or remain within the job market (among other), or to indeed receive further education (online).

Within these few examples above, the realities brought forward by the 4IR then indeed seems to have several advantages, and indeed be almost very good. It seems unlikely then that any such circumstance could be perceived as causing or attracting anything almost very bad.

## **3. Some things, almost very bad**

With aspects such as the ease of access to online technologies and a more digital world, come other realities. On the one side the security realisms that, for instance, the utilisation of cloud computing effects and on the other hand, the concurrent security concerns and apprehension that the exponential increase in virtual lifestyles presents.

### **3.1 Technology advancements and privacy**

One has to consider matters such as data privacy and/or the lack/or leaking thereof, especially coupled with advancements such artificial intelligence, robotics and machine learning - and the plethora of laws, regulations and directives that are echoing these matters globally. Onik, Kim and Yang (2019) referred to this and stated that “Artificial Intelligence is at the top of this privacy leaking list” and stress that “Several researchers have identified the privacy risk of AI”. They go further in noting that “Real-time image processing reveals human

identity and leaks millions of personal information” and report their study to have found “major issues of AI and Robotics in perspective data privacy”, namely:

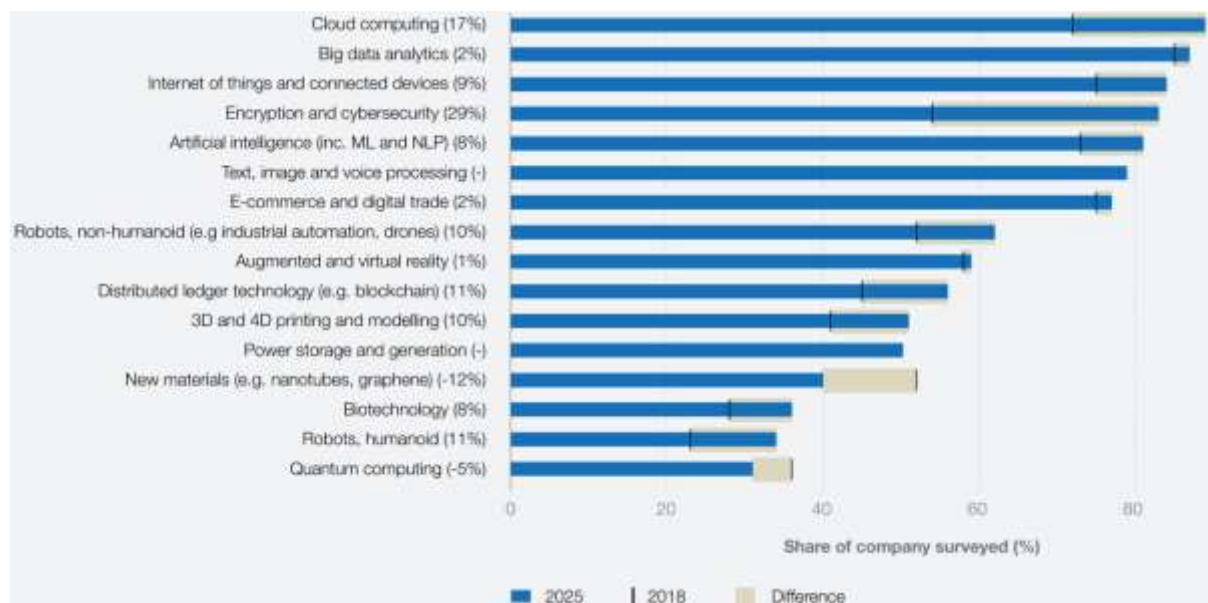
- “No privacy standardization for AI-based technologies”
- “Consent gathering from the user is inefficient”
- “AI decision making (profiling) should be monitored”

It then seems that technologies such as AI are advancing far quicker than the advancement in data protection policies, legislation and the development/implementation rate protection, monitoring and related defensive measures. It appears that the need for information, which drives business (and for that matter, which could also drive crime) could outweigh, or has the tendency of treating the need for privacy as an (although noted as important) afterthought.

### 3.2 Changes in technology uptake and job roles

Schwab (2015) interposes the “contradictions which shaped and continue to shape the so-called coming 4th IR”. He foresees that “We stand on the brink of a technological revolution that will fundamentally alter the way we live, work, and relate to one another”. He notes the intricacies of the 4IR and states that “In its scale, scope, and complexity, the transformation will be unlike anything humankind has experienced before”. He further declares that “We do not yet know just how it will unfold, but one thing is clear: the response to it must be integrated and comprehensive, involving all stakeholders of the global polity, from the public and private sectors to academia and civil society”.

In this regard the World Economic Forum references in their study (WEF, 2020) the marked difference in the “Technologies likely to be adopted by 2025”, as depicted below:



Source WEF, 2020

It could then also be argued that the same things we noted within section 2 above as being relatively good (for some), also has the potential of being relatively bad (for others). This technological revolution could have the effect that employees who were critical within an organisations value chain, could find that their skillsets are suddenly obsolete, especially if they do not have the resilience to keep their skill sets relevant (refer to section 5.3).

The same study from the WEF references the foreseen “Top 20 job roles in increasing and decreasing demand across industries” as follows:

↗ Increasing demand		↘ Decreasing demand	
1	Data Analysts and Scientists	1	Data Entry Clerks
2	AI and Machine Learning Specialists	2	Administrative and Executive Secretaries
3	Big Data Specialists	3	Accounting, Bookkeeping and Payroll Clerks
4	Digital Marketing and Strategy Specialists	4	Accountants and Auditors
5	Process Automation Specialists	5	Assembly and Factory Workers
6	Business Development Professionals	6	Business Services and Administration Managers
7	Digital Transformation Specialists	7	Client Information and Customer Service Workers
8	Information Security Analysts	8	General and Operations Managers
9	Software and Applications Developers	9	Mechanics and Machinery Repairers
10	Internet of Things Specialists	10	Material-Recording and Stock-Keeping Clerks
11	Project Managers	11	Financial Analysts
12	Business Services and Administration Managers	12	Postal Service Clerks
13	Database and Network Professionals	13	Sales Rep., Wholesale and Manuf., Tech. and Sci. Products
14	Robotics Engineers	14	Relationship Managers
15	Strategic Advisors	15	Bank Tellers and Related Clerks
16	Management and Organization Analysts	16	Door-To-Door Sales, News and Street Vendors
17	FinTech Engineers	17	Electronics and Telecoms Installers and Repairers
18	Mechanics and Machinery Repairers	18	Human Resources Specialists
19	Organizational Development Specialists	19	Training and Development Specialists
20	Risk Management Specialists	20	Construction Laborers

Source WEF, 2020

comprehensive study by the WEF assists in the validation of the perceived rapid move towards a digitised and virtual world, and indicates that it is foreseen that that this situation will rather intensify, than decline.

#### **4. Some, almost very ugly consequences**

When noting all these rapid changes (such as technology advancements, changes in technology uptake, privacy concerns and changing job roles); it then warrants interrogation around the opportunities that this situation presents to the cyber adversaries (whether it be a lone wolf, private enterprise, state sponsored or anywhere in-between) leading to potential very ugly consequences. It also warrants interrogation around the readiness of the cyber warrior community to, not just combat this, but to set or dictate the playing field with Initiatives to address these potential Ugly Consequences.

In as far back as 2011, Betz & Stevens, indicated that “cyberspace does have a direct effect on the information environment”. They further stressed the fact that it “is very disruptive of many processes heretofore considered safe such as the exchange of money and the relative security of personal, industrial and governmental data, as we can see from the burgeoning statistics on cyber-crime and cyber-espionage’

This sentiment is echoed by Stoddart (2016) indicating that “The volume, types and complexity of cybercrime, cyber espionage, and the kinds of APTs now being seen pose a problem that is not going to change unless more robust measures are put in place”.

Stadler (2020) indicates that “our reliance on technology and consumer connectedness, coupled with rapid growth in the aggregation and liquidity of personalized data, has made us more vulnerable to cybercrime victimization and the malicious use of private data”.

Cybercrime is growing at alarming rate. It has already reached critical levels and it is estimated that by 2025, the global cost of cybercrime can be up to \$10.5 trillion USD annually (GlobeNewswire, 2020). This estimated statistic suggests that, from a defensive point of view, we are doing something wrong. There seems to be a disjuncture between the apparent capabilities of the malicious cyber ‘community ‘and its ability to capitalise on the opportunities as presented by the 4IR, versus the capabilities of, and/or strategies deployed by the cyber warrior community – and although there are different degrees of severity, this seems to be a global scenario.

## **5. Some cybersecurity governance imperatives and the 4IR**

There are numerous writings on governance, best practice and strategy, providing well-structured approaches to cybersecurity. Whatever the methodology, it is clear that a holistic approach to the concept of cybersecurity in securing the 4IR is needed, with specific mention to the protection of information and data management.

In the end, it is all about managing risk. The ITU (2018) states that entities should be encouraged to “prioritise their cybersecurity investments and to proactively manage risk”. This is easier said than done, especially within developing countries where the recent realities with regard to the COVID19 pandemic had a notable impact on a governments’ fiscus as well as private sector business bottom-line and overall spending capability.

The ITU (2018) further states that prioritisation on cybersecurity investment should be prioritised “depending on an entity’s risk appetite” and notes that “a balance has to be maintained between security measures and potential benefits, considering the dynamic nature of the digital environment”. Although this statement is a best practice concept, this prioritisation on investment will be increasingly challenging as the move to the digital existence becomes the norm. Especially when this move necessitates the requirement for increasing cybersecurity expenditure in line with the need of the 4IR, whilst available funding, especially within developing countries remains under increasing strain.

With this in mind, there are three areas that, in the mind of the authors, that would be the most beneficial areas of focus as part of governance imperatives for securing cyber in the 4<sup>th</sup> Industrial revolution:

### **5.1 Strategy**

Implementation without strategy leaves you open to perpetual change driven by, among other sporadic events, changes in the prioritisation of funding and perpetual fluctuation in business focus and priorities. Proper strategic planning encourages appropriate implementation and assists to ensure that a holistic approach is followed in the identification and implementation of cyber primacies - in line with identified risk management requirements. Proper strategic planning also ensures that the cybersecurity focus is in line with the long term business strategy and as such, safeguards strategic ownership and sustainability.

This is resonated by Strategy& (2014), (formerly known as Booz & Company), indicating that “a better approach is to adopt a holistic view of information and technology, by considering all aspects of information superiority, doctrine, processes, people, training, and equipment. This approach looks at the entire life cycle of information, and it puts strategy at the forefront”. Strategy& (2014) go a step further by referring to “Specifically, developing an information superiority capability requires following five imperatives”:

- “treating information as a strategic asset”
- “having centralised governance”
- “building an information culture”
- “taking the right cyber security posture”
- “designing and delivering an integrated ICT infrastructure”

The exponential growth in cybercrime (as noted in section 4) compels countries to encompass cybersecurity as an essential part of their strategic plans (usually resulting in a National Cybersecurity Strategy). It is also well documented that one of the core elements of a good cybersecurity strategy is to ensure that a country has the relevant levels of cybersecurity knowledge and capabilities at all levels within the country (see Building Capacity in Section 5.3), as well as effective cyber intelligence and cyber counterintelligence capabilities. Unlike governments and large corporates, smaller role players would not always have sufficient resources for a solely dedicated intelligence and counterintelligence capacity. Nonetheless, intelligence and counterintelligence can, and ought to be part, of the overall organisational approach regardless of size (Jaquire, Duvenage & von Solms, 2018). This imperative is discussed in the next subsection.

### **5.2 Intelligence and counterintelligence**

Borum *et al* (2014), among other, added concepts such as “Intelligence forming Strategic decisions” stressing the “importance and role of strategic cyber intelligence to support risk-informed decision-making, ultimately

leading to improved objectives, policies, architectures and investments to advance a nation or organization's interests in the cyber domain". Observing on the role of intelligence and counterintelligence in proactive meeting 4IR challenges in future, De Loitte (2020) asserts: "The aspiration to protect everything will continue to recede in favour of focused efforts that anticipate, detect and disrupt active threats. This approach is more labour intensive and requires experience to carry out activities such as the collection of relevant threat intelligence, compromise assessments or deceptive operations. Therefore, counterintelligence will be used primarily to thwart advanced threats as more basic threats should be covered by cyber hygiene, automation and security by design."

Green (2016) adds to this by referring to further concepts and/or methods such as differential privacy, in which an "enhanced level of privacy protection in the evolving data security model" is achieved, "resulting in virtually no disclosure risk". Green (2016) explains that this is achieved "by obscuring individual identities with the addition of mathematical "noise" to particular data elements, consequently concealing a small sample of each individual's data". This can be combined with further deception and offensive techniques to identify adversaries and adversarial methodologies, or to even empower governments and/or private sector organisations to start to dictate the battle field, or to take the fight to the adversaries doorstep - if that should be appropriate .

In this spirit, ÆMaxima (2017) echoes such intelligence and counterintelligence thinking, stating that "One of the largest examples of weaponized information is in the area of denial & deception... outwitting your opponent by making them believe in something that will compel them to commit resources to an area that there is no present value or threat". Vinnakota (2017), links this back to governance and strategy, and includes the need for "A cybernetic governance model based approach considering the multidisciplinary aspects for the governance of cybersecurity with an ability to adapt continuously to a changing environment".

Keeping ahead of your adversary will increase your ability to secure your cyber environment. This is a bold statement, but one that has proved invaluable since the concepts such as 'intelligence' and 'counterintelligence', were first theorised millennia ago. Being able to understand who your adversary is, or who your potential adversaries are, knowing what they want or what they could possibly want from you; and being able to choose and control the battlefield will (as history has proven) give one the ultimate upper hand in winning the 'war' (Duvenage, Jaquire, & von Solms, 2020). Of course none of these ideas or initiatives are possible without adequate capacity and available/appropriate skills.

### **5.3 Building capacity and skills development**

No approach to cybersecurity can be effective without adequate capacity and skills. The lack of sufficient capacity and relevant skills will reflect in all stages of the cybersecurity process, from governance, to strategy, to implementation and maturity.

De Bruin and von Solms (2015) embrace the importance of capacity building within their cybersecurity governance maturity model structure. The authors reflect on the criticality around capacity with regard to People, Processes and Technology throughout the entire cybersecurity process and references the discussion of Alan Kaplan on "the necessary elements that need to be developed in order to have an effective capacity" namely:

- Developing a conceptual framework
- Establishing and organizational attitude
- Developing a vision and strategy
- Developing an organizational structure and,
- Acquiring skills and resources

Justiniano (2017) agrees with this view, indicating that "Although cyberspace danger and uncertainty are omnipresent, increased capacity building can mitigate the maliciousness". The International Telecommunications Union (ITU), in its 'Guide to Developing a National Cybersecurity Strategy', emphasises "the challenges related to advancing cybersecurity capacity-building and awareness-raising among government entities, citizens, businesses and other organisations – crucial to enabling a country's digital economy" (ITU, 2018).



The building of capacity also includes the continuous development and re-development of the skills of existing cyber personnel. Just as new capacity need to be obtained, existing human capacity needs to stay relevant in updating their own skillset (see the discussion within section 3.2). Governments and organisations also need to ensure that the skillsets of their current cyber workforce remain relevant. If these entities do not properly research changes in trends, threats, needs and operating practices and actively develop its exiting cyber skillsets accordingly, they might find themselves stocked with an obsolete cyber workforce in the near future, if not already.

## **6. Conclusion**

The inevitability of the fourth industrial revolution will “transform” not only “the workplace”, but the world of human existence in general. Recent global events (such as the COVID 19 pandemic) assisted in accelerating the implementation of 4IR concepts, fuelling the rapid cyberfication of systems and services - bringing with it a renewed emphasis on the realities that virtual lifestyle create. This situation fashions actualities that varies from the almost very good to the almost very bad, and ultimately underscores some almost very ugly consequences.

The digital age seems to provide much needed options to the small, medium and large enterprises, as well as individuals alike when it comes to being able to work off-site, maintaining a digital lifestyle, doing business, managing risk and levelling the playing field.

The same digital age also generates several security uncertainties and apprehension that the exponential increase in virtual lifestyles present. Issues such as technology advancements and its impact on privacy and/or the security of information, as well as the foreseen shift in technology uptake and the accompanied foreseen shift in job roles, could ultimately off-set the almost very good benefits, with almost very bad realities.

Recent events, such as the global outbreak of the COVID19 pandemic, had a notable impact on governments’ fiscus as well as the bottom-line and overall spending capability of private sector business, impacting the ability to implement best practice concepts such as the prioritisation of cybersecurity investments to proactively manage an organisations specific/unique risk.

The noted global increase in cybercrime merits further interrogation around the ability of cyber adversaries to capitalise on the (bad) cyber opportunities that the 4IR presents, vis-à-vis the readiness of the cyber warrior community to effectively counter this.

Although not the alpha and omega, there are three spheres which the authors identified that would be most beneficial areas of focus as part of governance imperatives for securing cyber in the 4<sup>th</sup> Industrial revolution; namely the need for ‘strategy’, ‘intelligence and counterintelligence initiatives,’ as well as the ‘building of capacity’.

It is noted that implementation without strategy could lead to perpetual change, a lack of funding and unending fluctuation in business focus and priorities. It is also noted that proper strategic planning assists in ensuring a cybersecurity focus in line with long term business strategy, which safeguards strategic ownership and sustainability.

Keeping ahead of your adversary increase one’s capability to secure the cyber environment. The ability to understand who the adversary is, knowing what they want and being able to choose and control the battlefield will ultimately assist in providing the upper hand to winning the ‘war’.

Cybersecurity initiatives cannot be effective without adequate capacity and skills, the lack of which will reflect in all stages of the cybersecurity process, from governance, to strategy, to implementation and maturity.

Apart from building new capacity, governments and private sector organisations need to ensure that the skillsets of their current cyber workforce remain relevant, or take the risk of finding themselves stocked with an obsolete cyber workforce – either now, or in the near future.

## References:

- ÆMaxima E, 2017, Information Supremacy Is the New Frontier of Combat, <https://medium.com/@emilymaxima/information-supremacy-is-the-new-frontier-of-combat-2b3e24936e55>, Accessed 07 February 2021
- Anser, M.K., Khan, M.A., Awan, U., Batool, R., Zaman, K., Imran, M., Sasmoko, Indrianti, Y., Khan, A., Bakar, Z.A., 2020. The role of technological innovation in a dynamic model of the environmental supply chain curve: evidence from a panel of 102 countries. *Processes* 8 (9), 1033. <https://doi.org/10.3390/pr8091033>
- Betz D & Stevens T, 2011, *Cyberspace and the State - Towards a Strategy for Cyber-Power*, 1<sup>st</sup> Edition Routledge.
- Bigerna S *et al*, 2021, *Technological Forecasting & Social Change* 165 (2021) 120558, accessed 07 February 2021
- Borum R *et-al*, 2014, Strategic cyber intelligence, [www.emeraldinsight.com](http://www.emeraldinsight.com), *Information & Computer Security* Vol. 23 No. 3, 2015 pp. 317-332
- De Bruin R and Von Solms S, 2015, *Modelling Cyber Security Governance Maturity*, SSIT, IEEE International Symposium on Technology in Society (ISTAS) Proceedings, ISBN978-1-4799-8283-7.
- de Castro Sobrosa Neto *et-al*, 2020, The fourth industrial revolution and the coronavirus: a new era catalysed by a virus, *ELSEVIER, Research in Globalization* 2 (2020) 100024, <https://www.journals.elsevier.com/resglo>, accessed 04 February 2021.
- Deloitte. 2020b. The future of cyber. Retrieved from <https://www2.deloitte.com/global/en/pages/about-deloitte/articles/gx-future-of-cyber.html>. Accessed 02 February 2021.
- Duvenage, P, Jaquire, V & von Solms, S, 2020, 'A Cyber Counterintelligence Matrix for Outsmarting Your Adversaries' in *Journal of Information Warfare*, 19(1): pp 1-11
- GlobeNewswire, 2020, [https://www.globenewswire.com/news-release/2020/11/18/2129432/0/en/Cybercrime-To-Cost-The-World-10-5-Trillion-Annually-By-2025.html#:~:text=Cybercrime%20To%20Cost%20The%20World%20%2410.5%20Trillion%20Annually%20By%202025,-very%20U.S.%20business&text=18%2C%202020%20\(GLOBE%20NEWSWIRE\),%243%20trillion%20USD%20in%202015](https://www.globenewswire.com/news-release/2020/11/18/2129432/0/en/Cybercrime-To-Cost-The-World-10-5-Trillion-Annually-By-2025.html#:~:text=Cybercrime%20To%20Cost%20The%20World%20%2410.5%20Trillion%20Annually%20By%202025,-very%20U.S.%20business&text=18%2C%202020%20(GLOBE%20NEWSWIRE),%243%20trillion%20USD%20in%202015), accessed 21 February 2021.
- Green 2016. What is Differential Privacy?—A Few Thoughts on Cryptographic Engineering, <https://blog.cryptographyengineering.com/2016/06/15/what-is-differential-privacy>, accessed 20 February 2020.
- ITU, 2018 <https://www.itu.int/myitu/-/media/Publications/2018-Publications/BDT-2018/Guide-to-developing-a-national-cybersecurity-strategy---Strategic-engagement-in-cybersecurity.pdf>
- Jaquire, V, Duvenage, P, & von Solms, S, 2018, 'Building the CCI dream team' in Published Proceedings of the 17th European Conference on Cyber Warfare and Security, Oslo, Norway, June.
- Justiniano E, 2017, *Advancing the Capacity of a Theater Special Operations Command (TSOC) To Counter Hybrid Warfare Threats In The Cyber Gray Zone*, Utica College, ProQuest Number: 10269999
- Kaplan A, 2015, Capacity building: Shifting the paradigms of practice, *Development in Practice*. 2000 08/01; 2015/06;10(3-4):517-26
- Muriel-Pera, *et al*, 2018, "Adoption of strategies the fourth industrial revolution by micro, small and medium enterprises in Bogota D.C.", Universidad Católica de Colombia
- Onik, Kim and Yang, 2019, Personal Data Privacy Challenges of the Fourth Industrial Revolution, *International Conference on Advanced Communications Technology (ICTACT)*, ISBN979-11-88428-02-1
- Penker M and Khoh S, 2018, *Cultivating Growth and Radical Innovation Success in the Fourth Industrial Revolution with Big Data Analytics*, 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 10.1109/IEEM.2018.8607313
- Sander, 2019, *The Sound of Silence: International Law and the Governance of Peacetime Cyber Operations*, 11th International Conference on Cyber Conflict, NATO CCD COE Publications, Tallinn
- Scepanović, 2019, *The Fourth Industrial Revolution and Education*, 8th MEDITERRANEAN CONFERENCE ON EMBEDDED COMPUTING, (MECO), 10-14 JUNE 2019, BUDVA, MONTENEGRO
- Schwab K, 2015, *The fourth industrial revolution: What it means and how to respond*, <https://www.foreignaffairs.com/articles/2015-12-12/fourth-industrial-revolution>, accessed 07 February 2021
- Stadler 2020, *Risks of Privacy-Enhancing Technologies: Complexity and Implications of Differential Privacy in the Context of Cybercrime*, Intechopen, DOI: 10.5772/intechopen.92752
- Stoddart K, 2016, *UK cyber security and critical national infrastructure protection*, The Royal Institute of International Affairs, *International Affairs* 92: 5
- Strategy&, 2014, *Achieving information superiority - Five imperatives for military transformation*, PWC, [www.pwc.com/structure](http://www.pwc.com/structure)
- Vinnakota T, 2016, *A second order cybernetic model for governance of cyber security in Enterprises*, IEEE DOI 10.1109/IACC.2016.136 696 695, IEEE6th International Conference on Advanced Computing.
- WEF, 2020, *The Future of Jobs Report 2020* <https://www.weforum.org/reports/the-future-of-jobs-report-2020>, accessed 04 February 2021.

# Critical Infrastructure Protection: Employer Expectations for Cyber Security Education in Finland

Janne Jaurimaa, Karo Saharinen and Sampo Kotikoski  
JAMK University of Applied Sciences, Jyväskylä, Finland

[m1270@student.jamk.fi](mailto:m1270@student.jamk.fi)

[karo.saharinen@jamk.fi](mailto:karo.saharinen@jamk.fi)

[sampo.kotikoski@jamk.fi](mailto:sampo.kotikoski@jamk.fi)

DOI: 10.34190/EWS.21.015

**Abstract:** In the human factor of cyber security, high level technical experts are considered as multidisciplinary technical gurus who are familiar with every aspect of IT environments including operating systems, code languages and protocols. University curricula and guiding frameworks, such e.g. NICE Cyber Security Workforce Framework, are designed to produce professionals to match the endless needs of working life. The cornerstones of achieving good working results can be considered as the level of expertise competence of the employee performing the task, as well as combining personal skills and abilities with the competence profile of the given task. Does the cyber domain need slightly lower educated, vocational level employees? As part of the National Security Policy in Finland, the vocational qualification in information and communications technology has recently started to produce suitable workforce for cyber labor on the European Qualifications Framework level 4 (EQF-4). In this research paper we answer the question how well the vocational education meets the demands of the employers as suitable workforce in cyber security in Finland. The study also investigated what kind of cyber security employees the Finnish employers currently need; what is the required level of education, level of experience and direction of competence. The research data was collected through a structured questionnaire survey, which was directed to critical national infrastructure protection companies such as Finnish telecom operators, ICT service providers, defense sector, and other governmental actors. The questionnaire results were examined with quantitative methods. Based on our results, regarding the content of education at EQF4-level, employers believe that the emphasis should be placed on basic technical skills and adherence to guidelines, while choosing more detailed specific areas of expertise is less important at this level of education. Based on the responses, in general cyber security related work has higher education level requirements than EQF4-level could provide. The results of the study can be used as guidelines for the development of the future curricula and in the strategic leadership of companies employing cyber security professionals.

**Keywords:** human factor, security policy, critical infrastructure protection, strategic leadership

---

## 1. Introduction

Finnish Cyber Security Strategy was published on 24 January 2013 in the form of a government resolution (The Security Committee of Finland, 2013). It specifies the main goals and operation models to meet the challenges in the cyber domain and ensure its functionality. In this first version, strategy is mentioned: *“The study of basic cyber security skills must be included at all levels of education”* and in the update it is stated that all cyber and information and communications technology (ICT) related training programs, including vocational level, will be strengthened (The Security Committee of Finland, 2019).

The EQF is an eight level framework which is designed to facilitate the comparison of national qualifications between EU countries (European Union, 2017). Finnish vocational qualification has been placed in level 4 of the EQF. The updated curriculum of Finnish Vocational Qualification in Information and Communications Technology introduced in August 2020 consists 180 competence points (Finnish National Agency for Education, 2020). The vocational qualification program graduates' students with five different qualification titles. In all of them, the module related to maintaining cyber security can be selected as an optional module.

The National Institute of Standards and Technology (NIST) has been the executor of National Initiative for Cybersecurity Education (NICE) in cooperation with the industry, government, and academia in the United States. Since its establishment in 2010, NICE has developed a working document draft of the NICE Cybersecurity Workforce Framework (NCWF), and in August 2017 it was published as NIST Special Publication 800-181 (NICE, 2017). The Framework is created to categorize and describe cyber security related work roles and tasks. It is designed to support many different parties including employers, employees, students, educators and technology providers. The framework provides a common lexicon as well as a taxonomy for the cyber security organizations and the workforce regardless of where or for whom the work is done.

At the highest level of the framework, cyber security work is then divided into seven categories. Inside the categories, there are 33 separate areas of cyber security work are called specialty areas. Each of them illustrates concentrated work or function in cyber security. The specialty areas contain 52 groupings called work roles, which consist of a set of specific knowledge, skills, and abilities (KSA) required to accomplish different tasks.

To create easier comparability for future researchers globally, in this research, the Finnish vocational qualification titles are converted to match the nearest corresponding NCFW work roles. For the Software Developer qualification title, a work role with the same name and similar work role was found in Securely Provision category. In the Operate and Maintain category two suitable work roles were found: qualification title pairs Network Operations Specialist-Networks Installers and Technical Support Specialist-IT Support Specialist. The mapping used in this research between the NICE framework and the Finnish vocational education can be illustrated in Figure 1.

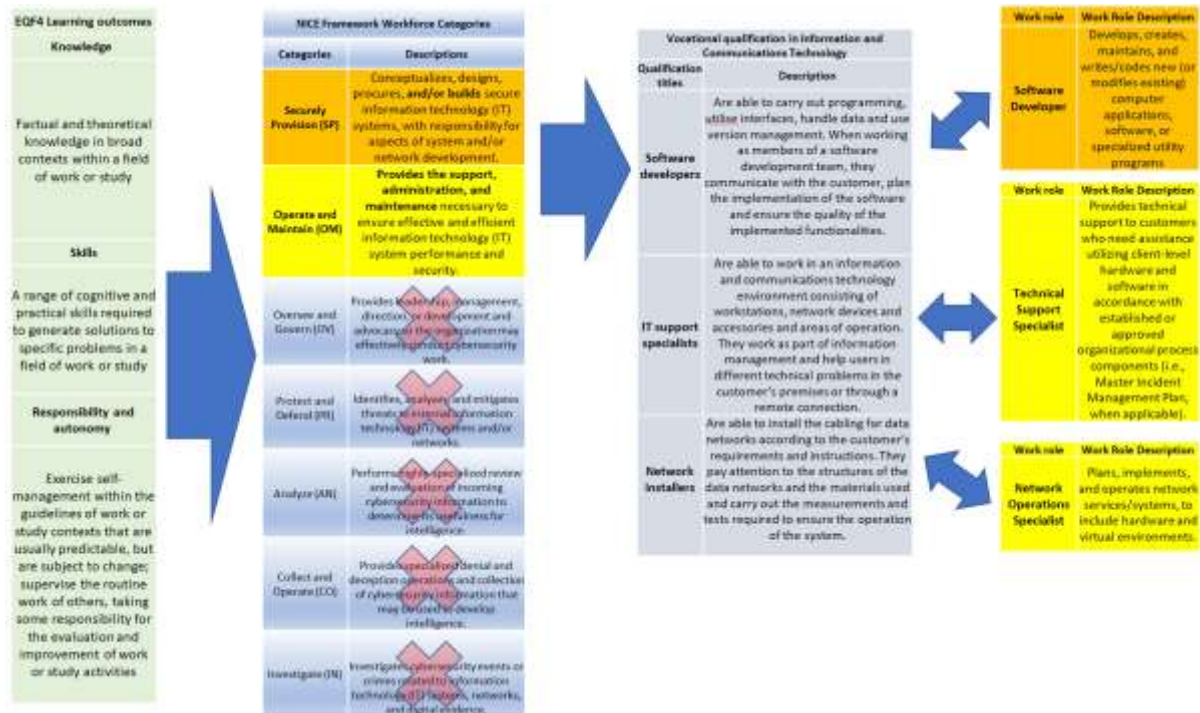


Figure 1: NCFW work roles and vocational qualification titles

## 2. Earlier research

The NICE framework has been used to develop degree programme structuration through *A Design Model for a Degree Programme in Cyber Security* to provide better targeted work role education for students on the Master's and Bachelor's Degree (Saharinen K., Karjalainen M., Kokkonen T., 2019). The emphases of different quantitative specialty areas have been researched regarding degree programme structuration (Backlund, J., 2020). The NCFW framework was utilized by matching the courses in curricula with the main categories of the framework. The research focused on the university level in the EU and the United States. The research contains a section on how the stakeholder demands between the industry and university education match one another.

Further influence was found in Jyväskylä Educational Consortium researched on the need for cyber security education in 2016 concentrating on Central Finland's SMEs. Simultaneously, also the teachers' perceptions of cyber security and cyber training were researched (Nevala, J., 2018). Similarly, the *Current and future needs of the cyber expertise in public sector organizations* publication researched two public sector organizations and their needs for cyber professional expertise through NCFW framework (Willberg, N., 2017).

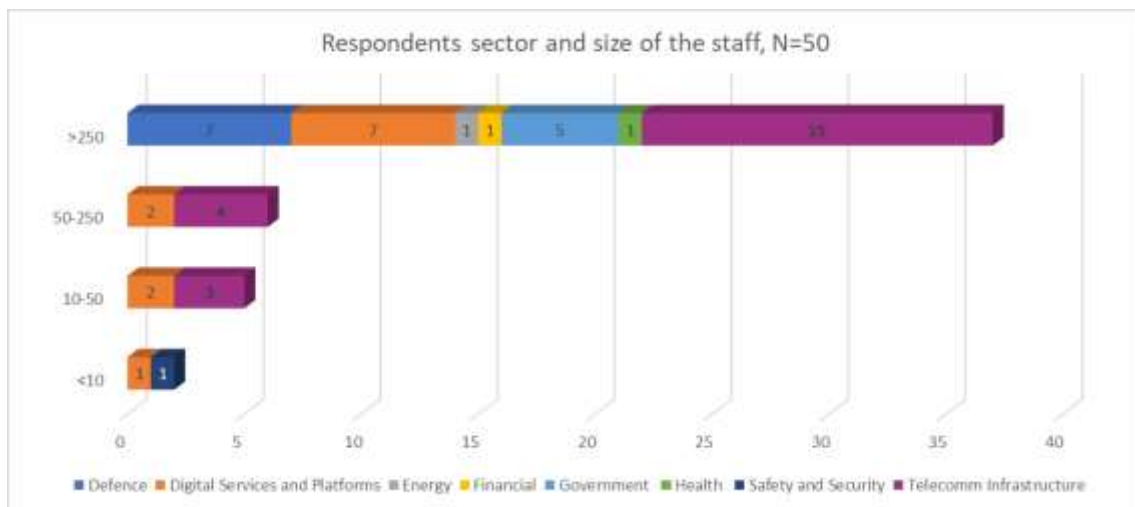
*Demand, availability and development of the cyber security workforce respond to the need for labor in Finland* examined the availability of cyber work in Finland from the recruiting organizations' point of view (Niemelä, J., 2019). In the study, the profile of a cyber professional employee is formed according to the requirements collected from employers. Cyber education in Finland is also evaluated focusing mainly on the universities and the universities of applied sciences.

As seen in the aforementioned paragraphs, there has been previous research on comparison of cyber education and frameworks with labor availability; however, the focus has not been on Finnish vocational education curricula or “apply” level workforce needs. This paper focuses its research on these sections, answering the question: Is vocational level cyber security education necessary as mandated in the Finnish Cyber Security Strategy?

### 3. Survey research from critical infrastructure the industry

The purpose of this survey was to find out the current suitability of Finnish cyber security education for different critical infrastructure industry in Finland. The main focus of the survey was on the Finnish Vocational Education (or qualification) in Information and Communications Technology. The survey also inquired and measured the importance of the education level and the amount of work experience required from the employer perspective in cyber related recruitment of jobs. Additionally, the labor needs for the cyber sector in Finland concerning near future were inquired about. As mentioned earlier, this research focuses on Network Installer, IT Support Specialist and Software Developer degree programmes and how necessary they are deemed.

The survey aimed at the organizations operating in Finland, which were classified according to sectorial division of the Proposal for a European Cybersecurity Taxonomy (JRC, 2019). The personnel size classification of companies is derived from an EU publication: The new SME definition (Publications Office of the EU, 2005). These commonly used classifications were used in the research to allow comparison with potential future research on the same kind of topic. The survey was conducted anonymously. The questions in the survey were implemented using a structured model to gather quantitative data. The questionnaire survey was active between 10 June 2020 - 26 July 2020 and it received a total of 50 responses. The responses came mainly from telecom operators, ICT service providers, defense sector and other governmental actors. The largest group of respondents were the large enterprises, which employ more than 250 employees. A sufficient number of Finnish actors in the field of critical infrastructure protection was involved. Figure 2 demonstrates the quantitative division of the respondents.



**Figure 2:** Respondents’ sector and size of the staff

### 4. Results

The respondents were asked to classify the qualification requirements of the cyber security maintenance related module of the curriculum of vocational qualification in Information and Communications Technology (ICT), according to importance of their business.

On a scale of 1-5, the mean of the responses was 3.96. Two modules even exceeded the 4.5 average, *Follows the information security instructions in their work* was considered the most important topic with 4.59 result, and the second most important topic was *Protects device with updates and software* with 4.51. More than four averages were also reached by topics *Makes development proposals to improve cyber security* 4.22 and *Monitor the data network using a variety of analysis tools* 4.04. Based on the responses, *Compares different encryption methods and selects the appropriate encryption method* 3.37 and *Scans vulnerabilities in the agreed network under review* 3.69 were considered as less important sections. In summary, citing the results it can be stated

that respondent organizations highly appreciate that at this level of education daily basic cyber security functions are carried out in accordance with the instructions.

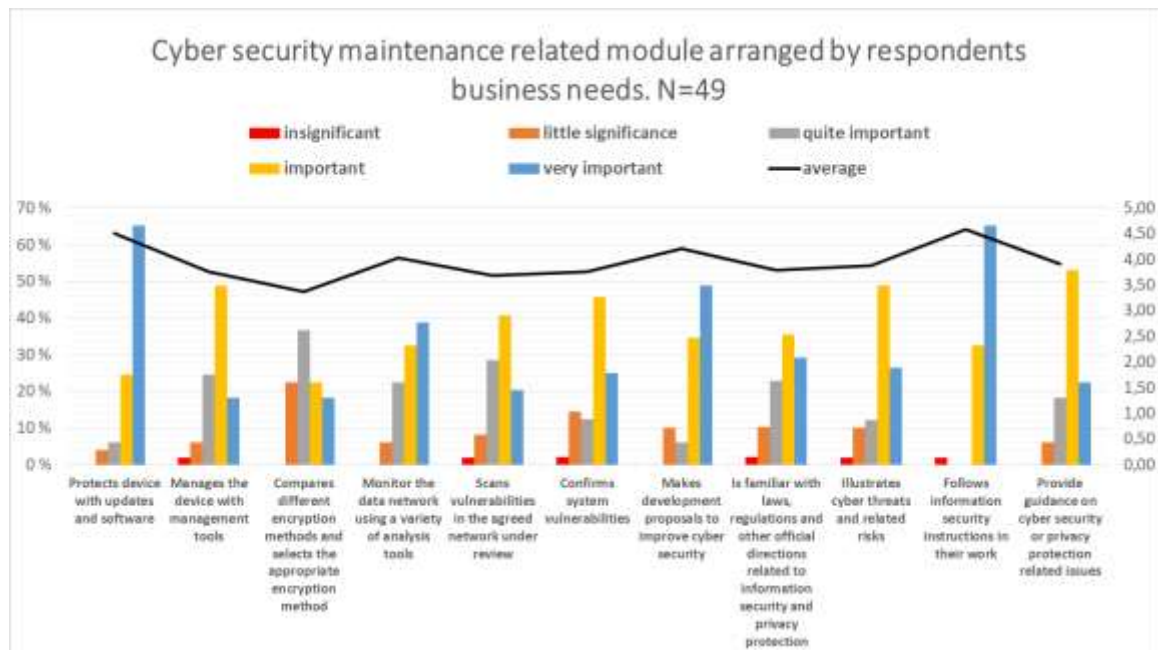


Figure 3: EQF4 Cyber security maintenance related module

The respondents were asked to classify the relevance of the cyber security modules in JAMK University of Applied Sciences' Information and Communication Technology degree program according to their importance to their business. The following Figure 4 demonstrates this distribution of answers.

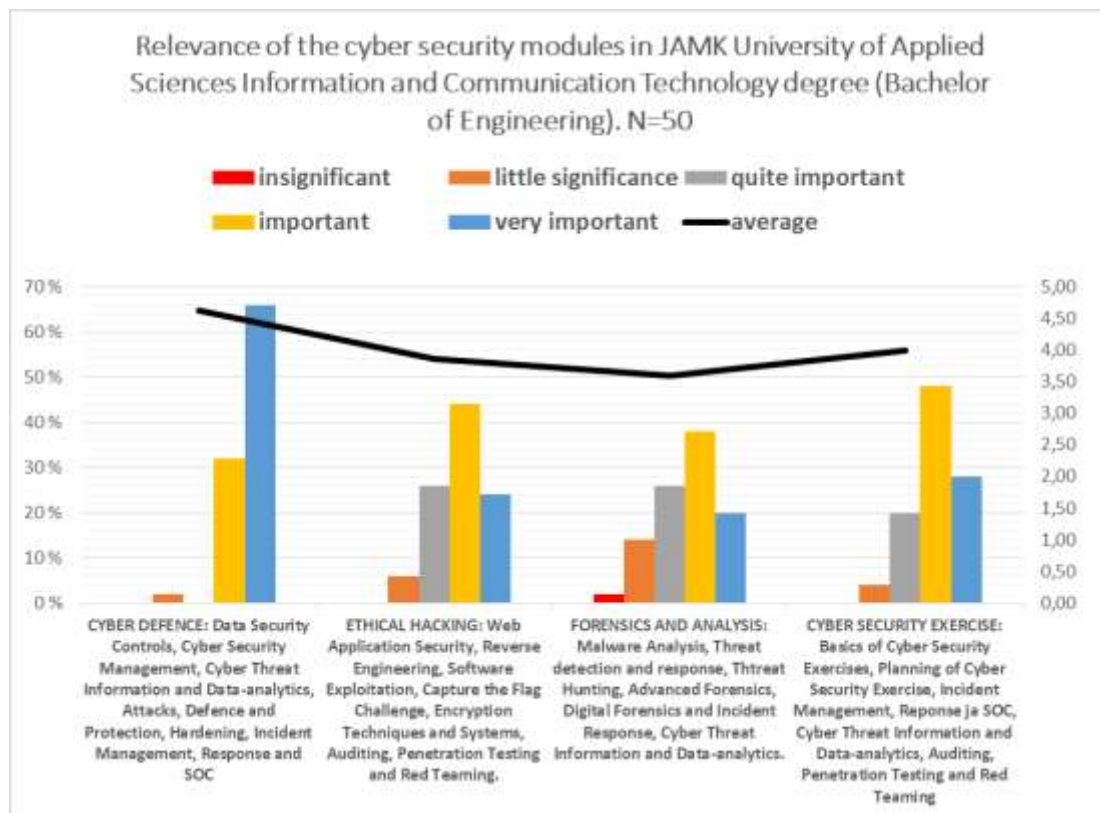


Figure 4: EQF6 Cyber security modules

Based on the responses, the same trend as earlier can also be seen in the content of the EQF6-level cyber security related modules; the modules are broader in content than at the EQF4 level, but there are fewer of them. In this



section on a same scale of 1-5, the mean of the responses was 4.02. One module exceeded the 4.5 average: *Cyber defence* was clearly considered as the most important topic with a result of 4.62, and the second most important topic was *Cyber security exercise* with 4.00. *Ethical Hacking* with a 3.86 result and *Forensics and analysis* 3.60 were considered as less important sections. According to the responses, fundamental knowledge of the cyber branch and practical hands-on doing seem to be important, and parts where more in-depth expertise is needed, are seen less relevant at this education level.

The respondents were asked about the near future labor needs of the selected work roles with EQF4-level experience. In this section, Finnish vocational qualification titles are converted to match the nearest corresponding work role in the NICE Cyber Security Workforce Framework work role. The following sample of results is seen in Figure 5: Near future work role needs for the EQF4-level experience

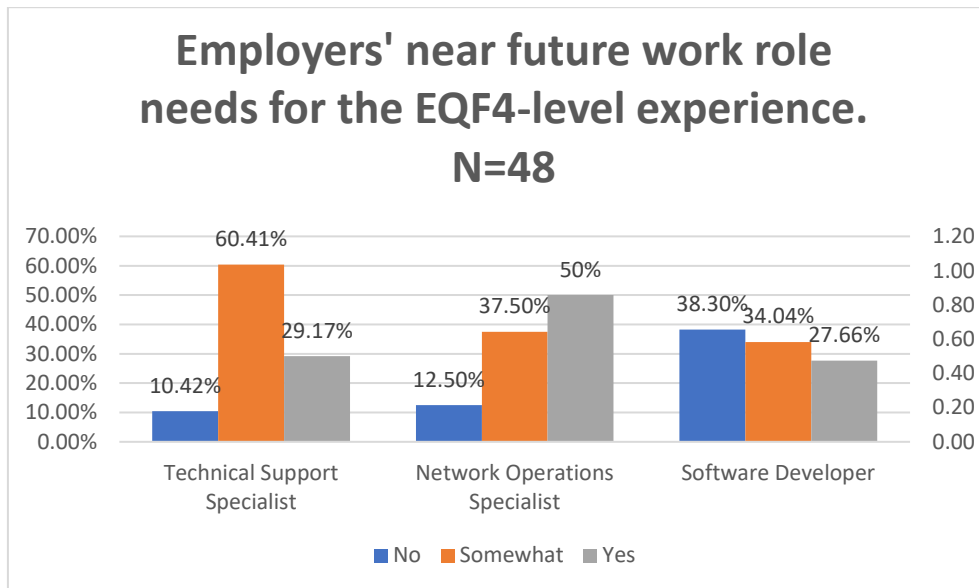


Figure 5: Near future work role needs for the EQF4-level experience

The greatest need for employees at this level of education is for network operations specialist and there may be a demand for technical support specialists. Software developers were the least needed at this level of education. High "Somewhat" bar might be explained by Technical support specialist role, which is often thought of as a helpdesk function, and many companies have outsourced this kind of role over the years. The respondents were asked about the target level of education when recruiting cybersecurity focused staff. In this question, it was possible to select multiple education level choices.

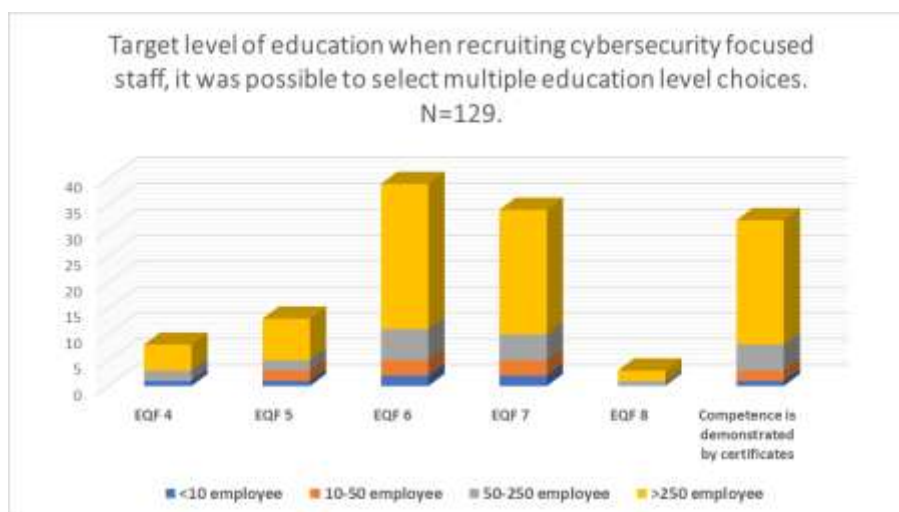


Figure 6: Target level of education

Based on the responses, cyber security related work in general have much higher education level requirements than EQF4-level could provide. University degrees and performed certificates are highly appreciated in the recruitment process. In addition to education level, another significant part in the selection process of the employee is the job applicant’s work experience. The respondents were asked about the needed level of experience when recruiting cyber security focused staff. The distribution of target experience levels for recruitment is shown in Figure 7: Target level of experience.



**Figure 7:** Target level of experience

The recruitment of entry-level employees is seen possible only in large companies. Career paths starting from the entry-level might be too challenging to smaller companies because they usually bind more experienced staff to the orientation process of a new entry-level employee. Based on the answers, intermediate experience level is the most popular class, but also the expert level is quite close to it. Lastly, the respondents were asked to assess the distribution of their company’s near future workforce needs based on the NCFW categories. The vocational qualification titles researched are divided into categories as follows: Securely Provision (SP) category includes vocational qualification title Software Developer. Operate and Maintain (OM) category includes titles Networks Installer and IT Support Specialist.



**Figure 8:** Near future workforce needs per NCFW category



The direction of the desired competence is strongly in Protect and Defend (PR) and Operate and Maintain (OM) categories; Securely Provision (SP) still fits in the top three categories. Analyze (AN) and Oversee and Govern (OV) clearly share the visions of respondent organizations; they both are seen necessary but also not necessary bar is high. Investigate (IN) and Collect and Operate (CO) categories are clearly seen as the least necessary.

Overall, from the employer's view, cyber security related subjects are seen as an important part of Information and Communication Technology education on both EQF-4 and EQF-6 levels. On a scale of 1-5, both modules were seen as averaging around four. Based on our research, on EQF-4 level it can be stated that the respondents consider compliance of their information security policies to be a very important part of their IT asset protection and expect this from every employee as well. Protection of devices with updates and software can be considered as one of the easiest ways to protect your environment against cyber threats, and this basic protection function seems to be appreciated by employers. Deviation notifications or any security development proposals are always valuable, especially if they are made proactively to mitigate potential threats. Situational awareness is again one of the basic functionalities of protecting an organization's most valuable data assets. Based on these results, the focus of education at this level should be on matched with basic security operations in accordance with the instructions, and more specific specializations should be given little less attention. For comparison, the same trend can be also seen in the content of the EQF-6 level cyber security related modules. The respondents' top rated module covers the basic techniques of cyber security field, and the module rated second goes through them in realistic hands-on exercise.

According to the responses, EQF-4 level education was not seen very appropriate for cyber related labor needs in Finland due to the higher level of education required for the cyber security focused staff. Overall, the chosen work roles were seen moderately appropriate. Generally, the greatest need for employees, out of the chosen degree specializations, at EQF-4 level of education is for Network Operations Specialist. Somewhat perhaps surprisingly, Software Developers were the least needed. Possibly the knowledge of basic techniques is valued more on this level of education, and the competence requirements of Software Developers are on a higher level in the surveyed organizations. The most suitable level of education when recruiting cybersecurity focused staff in Finland was EQF-6 and close to it was EQF-7; also the competence demonstrated in the certificates was considered appropriate.

The experience level of cyber security related employees is expected to be at least intermediate level; entry-level recruitment was only seen possible in two large companies. If the employee has got the ability to apply knowledge and skills in routine work situations without continuous guidance, the employee does productive work at least most of the time and does not appear as a mere expense during work induction. On the other hand, the expert level could be higher if there were enough qualified candidates available for the open cyber related vacancies.

The most needed direction of competence seems to be under Protect and Defend (PR) and Operate and Maintain (OM) categories. Identification, analysis, and mitigation of threats seem to be phenomena that responder companies still want to strengthen internally to have better cyber resilience. This research shows that they are willing to recruit their own employees to enhance the capability. Applications and devices are constantly evolving; hence admins must update and patch existing systems while new features or systems are introduced. They also want to keep these basic functions in their own hands, and an operator for these responsibilities would also be needed internally. These responses describing labor needs show a clear link to needs related to education priorities, strong basic knowledge of computing and information security, and practical hands-on skills are valuable. From the research data of the surveyed companies it can be concluded, that if they use more advanced cyber security services, like forensics, advanced analysis, or ethical hacking services, they might mainly outsource them to high-tech partners and do not recruit these employees themselves. This would explain the low demand for labor in these sectors.

## **5. Conclusion and future research**

Based on our research, the profile of most wanted cyber employees' direction of competence is strong system/network administrator who knows how to operate, maintain, and mitigate threats in the environment for which they are responsible, and they should have at least EQF-6 level education and a minimum of intermediate level work experience. The competences demonstrated with certificates were considered very important, so they can be seen as a significant part of professionalism also in the cyber field. An earlier research

published in 2019 by Jukka Niemelä states that there is a clear shortage of suitable labor in the cyber sector in Finland. The expected level of competence of the applicants has been lowered, and it is hoped that in the future applicants will have a basic knowledge of the cyber branch and deep expertise in one of the key areas of cyber security (Niemelä, J., 2019). On this basis, vocational qualification does not solve the problem encountered in the previous research, and in order to gain deep expertise further education or specialization in working life are still needed. As mentioned earlier, there are no open cyber related vacancies for EQF4-level graduates as inspected by the authors of this paper. However, vocational qualification gives a good starting point for vocational work tasks, as well as the keys for life lasting learning in further education and in career progression. Strong practical hands-on skills should be achieved during vocational training, whether they consist of network technology, programming, or different operating systems. If it is desired to steer career pathway from the basic ICT tasks to the direction of cyber security, the options are either to carry out industry certifications or accomplish further education. The aim of further education should be to deepen strong basic skills to the specialization in the chosen cyber expertise area.

For future research, it would be interesting to investigate how cyber security has been implemented in other countries “on all levels of education” as Finland’s cyber security strategy mandates. The qualitative research data also emphasized the ‘soft skills’ of sought out employees, not just the ‘hard technical skills’. It is also a debatable subject, where the subject of cyber security should be sectioned and emphasized as an own educational field, as many of the curriculum proposals currently entangle it along every subject. This could be investigated through the workforce demand for different levels of education.

## **Acknowledgements**

This work has been done in Jyväskylä University of Applied Science (JAMK) which is participating the Cyber Security Network of Competence Centres for Europe (CyberSec4Europe) project of the Horizon 2020 SU-ICT-03-2018 program. <https://cybersec4europe.eu/about/>

The authors would like to thank Tuula Kotikoski for her contribution in proofreading the English language on the paper.

## **References**

- Backlund, J. (2020) Examination of contemporary cyber security education [Online]. Available at <http://urn.fi/URN:NBN:fi:amk-2020060416851> [Accessed 25 August 2020].
- European Union (2017) Description of the eight EQF levels [Online]. Available at <https://europa.eu/europass/en/description-eight-efq-levels> [Accessed 5 September 2020].
- Finnish National Agency for Education (2020) Qualification requirements entered into force on 01.08.2020 (OPH-2596-2019) [Online]. Available at <https://eperusteet.opintopolku.fi/eperusteet-service/api/dokumentit/6941346> [Accessed 25 August 2020].
- Joint Research Centre (JRC), the European Commission’s science and knowledge service (2019) A Proposal for a European Cybersecurity Taxonomy [Online]. Available at <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC118089/taxonomy-v2.pdf> [Accessed 5 September 2020].
- National Initiative for Cybersecurity Education (2017) Cybersecurity Workforce Framework [Online]. Available at <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-181.pdf> [Accessed 1 August 2020].
- Nevala, J. (2018) Cybersecurity situation analysis - Survey in Central Finland 2016-2018 [Online]. Available at <http://urn.fi/URN:NBN:fi:amk-2018121721956> [Accessed 19 September 2020].
- Niemelä, J. (2019) Demand, availability and development of the cyber security workforce respond to the need for labor in Finland [Online]. Available at <http://urn.fi/URN:NBN:fi:jyu-201906032891> [Accessed 25 August 2020].
- Publications Office of the EU. (2005) The new SME definition [Online]. Available at <https://op.europa.eu/en/publication-detail/-/publication/10abc892-251c-4d41-aa2b-7fe1ad83818c> [Accessed 5 September 2020].
- Saharinen K., Karjalainen M., Kokkonen T., (2019) A design model for a degree programme in cyber security [Online]. Available at <https://doi.org/10.1145/3369255.3369266> [Accessed 25 August 2020].
- The Security Committee of Finland (2013) Finland's Cyber security Strategy [Online]. Available at [https://www.defmin.fi/files/2378/Finland\\_s\\_Cyber\\_Security\\_Strategy.pdf](https://www.defmin.fi/files/2378/Finland_s_Cyber_Security_Strategy.pdf) [Accessed 29 August 2020].
- The Security Committee of Finland (2019) Finland's Cyber security Strategy 2019 [Online]. Available at [https://turvallisuuskomitea.fi/wp-content/uploads/2019/10/Kyberturvallisuusstrategia\\_A4\\_ENG\\_WEB\\_031019.pdf](https://turvallisuuskomitea.fi/wp-content/uploads/2019/10/Kyberturvallisuusstrategia_A4_ENG_WEB_031019.pdf) [Accessed 29 August 2020].
- Willberg, N. (2017) Current and future needs of the cyber expertise in public sector organizations [Online]. Available at <http://urn.fi/URN:NBN:fi:jyu-201706243034> [Accessed 25 August 2020].

# Digital Forensic Readiness Implementation in SDN: Issues and Challenges

Nickson Karie<sup>1, 2</sup> and Craig Valli<sup>1, 2</sup>

<sup>1</sup>Cyber Security Cooperative Research Centre, Australia

<sup>2</sup>Security Research Institute, Edith Cowan University, Australia

[nickson.karie@cybersecuritycrc.org.au](mailto:nickson.karie@cybersecuritycrc.org.au)

[c.valli@ecu.edu.au](mailto:c.valli@ecu.edu.au)

DOI: 10.34190/EWS.21.091

**Abstract:** The continued evolution in computer network technologies has seen the introduction of new paradigms like Software Defined Networking (SDN) which has altered many traditional networking principles in today's business environments. SDN has brought about unprecedented change to the way organisations plan, develop, and enact their networking technology and infrastructure strategies. However, SDN does not only offer new opportunities and abilities for organisations to redesign their entire network infrastructure but also presents a different set of issues and challenges that need to be resolved. One such challenge is the implementation of Digital Forensic Readiness (DFR) in SDN environments. This paper, therefore, examines existing literature and highlights the different issues and challenges impacting the implementation of DFR in SDN. However, the paper also goes further to offer insights on the different countermeasures that organisations can embrace to enhance their ability to respond to cybersecurity incidents as well as help them in implementing DFR in SDN environments.

**Keywords:** digital forensic readiness, software defined networking, issues and challenges, cyber security incidents, countermeasures

---

## 1. Introduction

Software-defined networking (SDN) can be defined as “an approach to networking that uses software-based controllers or application programming interfaces to communicate with the underlying hardware infrastructure and direct traffic on a network” (VMware, 2021). According to Kandoi and Antikainen, (2015) SDN has recently gained significant momentum in many organisations. Amin, Reisslein, and Shah, (2018) adds that SDN gives organisations the power to decouple the control plane from the data plane of forwarding devices thus simplifying network management and control, making computer networks agile and flexible. However, Mugitama, Cahyani, and Sukamo (2020) state that since its inception, the SDN architecture which mainly abstracts different, distinguishable layers of a network was not designed with a focus on network security. For this reason, the centralized control nature of SDN introduces some security vulnerabilities that can be exploited, for example, using Denial of Service (DoS) attacks to cause packet overload or race condition on the controller.

Another known security vulnerability at the SDN controller level is the topology poisoning attack. This attack utilizes spoofed packets and exploits the Link Layer Discovery Protocol (LLDP) packets in the network. Topology poisoning attack is one among many other malicious activities and security vulnerabilities that makes Digital Forensic Investigation (DFI) a challenging process in SDN environments. Also, the lack of standardised approaches specifically designed to help forensic investigators in SDN environments adds to the complexity of conducting digital forensic investigations in SDN.

As organisations continue to embrace SDN, many of them are likely to be targeted or abused by malicious actors to facilitate malicious cyber activities (Kebande, et.al, 2020). The ability to execute malicious activities in SDN environments to cause harm or abuse makes SDN forensics an increasingly important process and reinforces the importance of implementing DFR in organisations (Karie and Karume, 2017).

This paper, therefore, examines existing literature and highlights the different issues and challenges impacting the implementation of DFR in SDN environments. Further, this paper highlights different countermeasures that organisations can embrace to enhance their ability to respond to cybersecurity incidents as well as help them in implementing DFR in SDN environments. Note at this point also that this paper utilised purposive sampling research methodology where the authors relied on their own judgment to choose the literature examined and used in this study.

As for the remaining part of this paper: Section 2 introduces the literature review as background while Section 3 presents related work. Thereafter, Section 4 discusses the issues and challenges impacting the implementation of DFR in SDN. Section 5 presents different countermeasures that organisations can embrace to enhance their ability to respond to cybersecurity incidents as well as help them in implementing DFR in SDN environments before concluding the paper in Section 6 and makes mention of the future work.

## **2. Background**

This section provides a literature background on the following areas: digital forensics, Digital Forensic Readiness (DFR), and Software-Defined Networking (SDN). Digital forensics helps to understand the scientific process used for conducting digital forensic investigations while DFR is discussed as a way that can help organisations record activities and data in such a manner that the records are sufficient in their extent for subsequent forensic purposes thus minimizing digital forensic investigation costs (Mohay, 2005). Finally, SDN is discussed to help understand the concept of physical separation of the network control plane from the forwarding plane and how the control plane is used to control different devices in the network.

### **2.1 Digital forensics**

Continued malicious activities in the cyberspace makes digital forensics an essential process for investigating and prosecuting cybercriminals misusing digital devices such as computer systems, network devices, mobile devices, and storage devices (Selamat, Yusof, and Sahib, 2008). For this reason, digital forensics can be described as a scientific process of investigation that deals with extracting and analysing digital artefacts from digital devices. According to Ademu, Imafidon, and Preston (2011), digital forensics provides tools, techniques, and scientifically proven methods that can be used to acquire and analyse digital artefacts. The acquired or extracted digital artefacts can be used to reconstruct events for purposes of creating a hypothesis that can be useful in a court of law or any civil proceedings (Karie and Kebande, 2016).

Any successful digital forensic investigation process involves a thorough forensic examination of digital artefacts by forensic analysts with the primary objective being to unearth information that will assist practitioners and law enforcement agencies in the presentation of digital evidence in a court of law or any civil proceedings. In this regard, Karie and Kebande (2016) add that digital forensics thus expands simply from the crime scene through digital forensic analysts to the courtroom. This also implies that, if the outcome of an investigation is to be presented in a court of law as evidence, the digital forensic investigation process must always adhere to some important scientifically proven and accepted methods or processes that must be considered and taken (Köhn, Olivier, and Eloff, 2006). For a comprehensive reading on digital forensic investigation processes, the reader is advised to consult Valjarevic, and Venter, (2015), Valjarevic and Venter, (2012), and Köhn, Olivier, and Eloff, (2006). The next subsection presents the concept of digital forensic readiness.

### **2.2 Digital Forensic Readiness (DFR)**

Based on the definitions and description of digital forensics, the digital forensic investigation process is usually employed as a post-event response to cybersecurity incidents (Rowlingson, 2004). However, for an organisation to benefit from the ability to gather and preserve digital evidence before an incident occurs, DFR is inevitable. DFR can be defined as “the ability of an organisation to maximize its potential to use digital evidence whilst minimizing the costs of an investigation” (Rowlingson, 2004). From this definition, DFR allows an organisation to regain control and limit the damage and costs from any security incident (KPMG, 2015) as well as demonstrate due diligence with regulations, conduct digital forensic investigations, and produce digital artefacts that can be used in a court of law or civil proceedings as evidence (DFMAG, 2020). This also implies that DFR can help organisations record activities and data in such a manner that the records are sufficient in their extent for subsequent forensic purposes thus minimizing digital forensic investigation costs (Mohay, 2005).

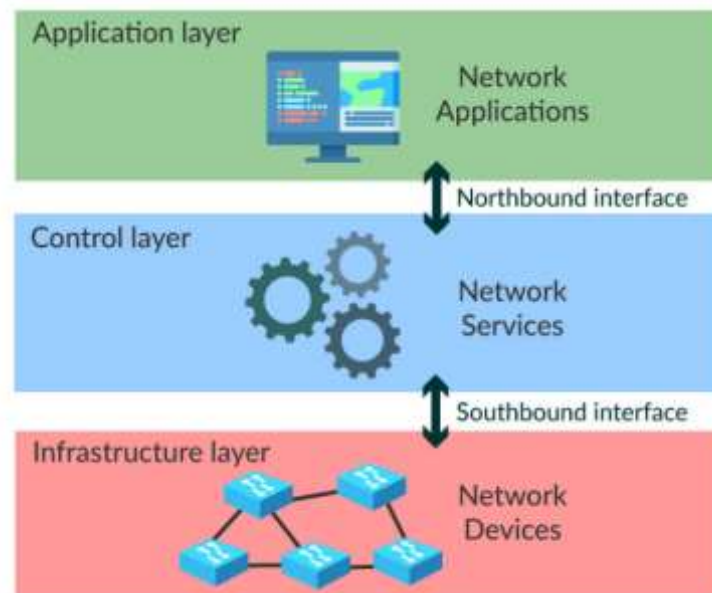
Antonio and Labuschagne (2013) add that a carefully considered and planned legally contextualized DFR strategy can provide organisations with an increased ability to respond to security incidents while maintaining the integrity of the evidence gathered and keeping investigative costs low. DFR can also help organisations with quicker recovery, improved business continuity, and compliance, as well as an improved success rate in legal actions by having available the collected digital artefacts for use as evidence (Andre, 2014). From this DFR description, the authors agree with the sentiments echoed by Elyas, Ahmad, Maynard, and Lonie, (2015) that

implementing DFR in SDN environments can help organisations to comply with their legal, contractual, regulatory, security, and operational obligations. The concept of SDN is covered in the next subsection.

### 2.3 Software-Defined Networking (SDN)

SDN is an emerging networking technology offering centralised control of computer networks (Mugitama, Cahyani, and Sukamo, 2020) that allows network operators to dynamically configure and manage their infrastructures. The primary difference between SDN and traditional networking is that SDN is software-based, while traditional networking is hardware-based (VMware, 2021). Because of the programmability of SDN, organisations can experience more advantages of having SDN than traditional networking techniques. Some of the primary advantages of SDN are increased control with greater speed and flexibility, customizable network infrastructure as well as robust security.

The goal of SDN however as stated by Rosencrance, English, and Burke, (2020) is to “improve network control by enabling enterprises and service providers to respond quickly to changing business requirements”. This is backed up by the fact that in SDN, “a network engineer or administrator can shape traffic from a centralized control console without having to touch individual switches in the network. Besides, a centralized SDN controller will direct the switches or routers to deliver network services wherever they are needed, regardless of the specific connections between a server and devices” (Rosencrance, English, and Burke, 2020). Figure 1 shows a simplified SDN architecture.



Source: (<https://electronicsguide4u.com/sdn-network-software-defined-network-openflow-protocol-what-is-sdn/>)

**Figure 1:** Software-Defined Network (SDN) architecture

Infer from Figure 1 that SDN can be divided into three-tier architecture with three primary layers. Communication through interfaces between the different tiers or layers is usually handled by the northbound and southbound interfaces as shown in Figure 1. The three basic layers of SDN are brief explained in the following subsections.

#### 2.3.1 Application layer

The application layer as shown in Figure 1 hosts the SDN applications or network applications which are programs designed to perform different tasks. SDN applications explicitly, directly, and programmatically communicate their network requirements and desired network behaviour to the SDN controller via a northbound interface (ONF, 2013). The northbound interface (NBI) in this case serves as a communication link between SDN applications and the SDN controller. Again, the NBI provides an abstract network view that enables the direct expression of network behaviour and requirements. Examples of SDN applications include networking management, analytics as well as other business applications used mostly in data centres (SDx, 2015).

### *2.3.2 Control layer*

The control layer oversees the network intelligence and hosts the control logic for managing the network services. The SDN controller which also acts as the brain of the entire network manages and manipulates flow entries to and from multiple devices (Aric, 2020). As seen from Figure 1 the control layer houses the SDN controller which is a logically centralized entity. According to ONF (2013), some of the functions of the control layer include but are not limited to translating the requirements from the application layer to the data paths and providing SDN applications with an abstract view of the network (ONF, 2013). Note that the SDN data path is a software-based representation of a network device in SDN environments (Kemp,2020) that exposes visibility and uncontested control over its advertised forwarding and data processing capabilities (ONF, 2013).

### *2.3.3 Infrastructure layer*

The infrastructure layer which also forms the physical layer of the network consists of various network devices (network switches and routers) or networking equipment (Aric, 2020) and forms the underlying network to forward network traffic. The network devices found in the infrastructure layer are responsible for handling network packets based on the rules provided by the SDN controller. The next section presents related works.

## **3. Related works**

Several related research work from different researchers exists and have made valuable contributions to the study presented in this paper. In this section, a summary of some of the most prominent efforts is presented.

To begin with, Munkhondya, Ikuesan, and Venter, (2019) proposed a DFR approach for potential evidence preservation in SDN. However, their research focused on developing a proactive mechanism for the identification, handling, collection, and preservation of digital artefacts in SDN. Also, their proposed mechanism was meant to integrate the DFR approach to the acquisition and preservation of volatile artefacts. Their research however did not capture the specific issues and challenges impacting the implementation of DFR in SDN environments as is the case of this current study.

Another effort by Lagrasse, Singh, Munkhondya, Ikuesan, and Venter, (2020) proposed “a proactive DFR framework for SDN with a trigger-based automated collection mechanism. Their proposed mechanism integrated an intrusion detection system and an SDN controller”. Their research mainly concentrated on how best to collect digital artefacts in SDN environments without regard to the issues and challenges impacting the implementation of DFR in SDN. The current study in this paper, however, investigates this research gap by discussing the issues and challenges impacting the implementation of DFR in SDN environments.

Sezer, et.al. (2013) raised the question of how to achieve a successful carrier-grade network with SDN. In their research, however, Sezer, et.al. (2013) only focus on the challenges of network performance, scalability, security, and interoperability with the proposal of potential solution directions. Their research had no reference to the issues and challenges impacting the implementation of DFR in SDN environments as is the case discussed in this current paper.

Efforts by Munkhondya, Ikuesan, and Venter, (2020) presented a case for a dynamic approach to DFR in SDN environments. Their research argues that “the DFR approach mostly employed has been limited to static potential digital evidence collection which contrasts the dynamic nature of SDN environments” (Munkhondya, Ikuesan, and Venter, 2020). They then go ahead and discuss the pitfalls of the static DFR approach which underscores the need for a dynamic DFR approach. The current paper however focuses on discussing the different issues and challenges impacting the implementation of DFR in SDN environments.

Research by Park, et al. (2018) designed a digital forensic readiness model for a cloud computing-based smart work environment considering the current changes in cloud computing. Their research work was purely based on a cloud computing-based smart work environment and not on issues and challenges impacting the implementation of DFR in SDN environments.

Spiekermann and Eggendorfer, (2016) analysed different challenges in investigating virtual networks. As opposed to the current study in this paper, their research proposed a classification in several different categories to help in developing new methods and possible solutions to simplify investigations in virtual network environments and not necessarily SDN. Karie and Karume, (2017) also presented different issues and challenges

surrounding the implementation of digital forensic readiness in organisations. Their research, however, was generic and did not point out any specific implementation of DFR in SDN environments. Though the focus of the current paper is on SDN, the authors acknowledge that some of the different sentiments discussed by Karie and Karume, (2017) are also applicable to SDN environments.

Other related works exist on issues and challenges surrounding DFR in SDN environments, however, neither those nor the cited references in this paper have presented the specific issues and challenges impacting the implementation of DFR in SDN environments in the way that is discussed in this paper. However, the authors acknowledge the fact that the previous research works have offered valuable insights toward the study in this paper. The next section presents a detailed discussion of the different issues and challenges impacting the implementation of DFR in SDN environments.

#### **4. Issues and challenges impacting DFR implementation in SDN**

In this section of the paper, the authors discuss some of the identified issues and challenges impacting DFR implementation in SDN environments. Note that as mentioned earlier this study employed purposeful sampling and for this reason, the issues and challenges identified in this section were only selected to facilitate this study based on the literature sampled by the authors and do not by any means constitute an exhaustive list. Besides, some of the issues and challenges discussed are native to SDN environments hence, more specific issues and challenges can and should be added as technology evolves.

##### **4.1 Lack of standards focusing on DFR implementation in SDN**

Like any other existing standard, the primary reason for having internationally recognised DFR standards is “to promote good practise methods and processes for forensic capture and investigation of digital artefacts” (ISO/IEC 27037, 2012). This also implies that standards are generally accepted as good, however, due to the relative newness of SDN infrastructure, as well as having recently gained significant momentum in many organisations (Kandoi and Antikainen, 2015) international standards that focus purely on DFR implementation in SDN environments are yet to be realised. The lack of international standards specifically focusing on DFR implementation in SDN is therefore a challenge that according to Khan et al. (2016), calls for a comprehensive forensic mechanism or standard to help in investigating the different forms of attacks in SDN environments as well as facilitate future forensic investigations.

##### **4.2 Lack of DFR implementation policies for SDN**

To achieve business objectives many organisations, make use of policies. In the context of a legacy network system, policies are a collection of rules, conditions, constraints, and settings defined by network administrators to govern the behaviours of network devices or designate who is authorised to connect to the network and the circumstances under which they can or cannot connect (Microsoft, 2020). In the SDN environment, a DFR implementation policy can thus be understood as a document that details the immediate procedures to be employed to support DFR in an organisation and any future forensic investigation of digital artefacts. Such a policy provides a systematic, standardised, and legal basis for the admissibility of digital evidence that may be required from a formal dispute or legal process (Karie and Karume, 2017). However, correctly setting up and enforcing DFR implementation policies in SDN may require accurate, fine-grained, and trusted information of user applications generating network traffic which may not be available upfront in new SDN environments. A Lack of DFR implementation policies in SDN environments, therefore, makes the development and implementation of any SDN DFR frameworks a very challenging process.

##### **4.3 Budget constraints for implementing DFR in SDN**

Knowing that SDN is still considered relatively new, and that security was not initially a key characteristic of the SDN architecture according to Khan et al. (2016), implementing DFR may be subject to budgetary or cost constraints which may not be known upfront. Moreover, as stated by Reddy, Venter, and Olivier, (2012), a DFR program consists of several activities that should be chosen and managed for the cost constraints and risk of the organisation. However, as organisations opt for cheaper options to be implemented in the place of DFR because of budget constraints, some of the alternatives implemented make digital forensic investigation costly and challenging.

#### **4.4 Fear of network downtime**

When operating on legacy network systems, which in most cases is used to support client business, migrating to SDN means a part of the entire legacy network will have to be shut down for some time causing client business needs to be shut down for the length of the specified downtime as well. This scenario may affect the implementation of DFR in SDN environments especially in an organisation where the process involved cannot be fully automated to facilitate the integration of the legacy network systems into SDN environments while maintaining backward compatibility (QuoteColo, 2016). The fear of network downtime, therefore, is a challenge when an organisation is considering a move to SDN.

#### **4.5 Lack of skilled personnel**

In many business environments, it is expected that the skills offered by the personnel match the skills wanted by organisations to meet their objectives. However, in most cases, this is always not the case. According to Desai et al. (2009), for some time now knowledgeable and skilled digital forensic personnel are hard to find. The shortage of skilled personnel poses a challenge in many different environments including the implementation of DFR in SDN. This, therefore, calls for organisations to address skill shortages to cover existing gaps through training or workshops which can also have other budget or cost implications.

#### **4.6 SDN scalability challenge**

As businesses grow, so is their networking environment. Large SDN environments with volumes of networking requests can overwhelm SDN controllers making the management of information flow between the separate data plane and control plane a challenge to SDN. One possible solution is for organisations to embrace decentralized control architecture as well as more intelligence implemented to the data control plane to intersperse data between multiple control planes. This will also help in monitoring and ensure network device accountability and network latency between connected planes. However, decentralized control architecture can make implementing DFR a challenge to some organisations.

#### **4.7 DFR implementation as a challenge**

Existing standards like the ISO/IEC 27043 are more generic and not application specific according to KEBANDE, MUDAU, IKUESAN, VENTER, and CHOO, (2020). Compared to legacy network systems, SDN has some level of complexity hence implement SDN technology without reinventing the whole architecture with its aspects and related components can be challenging according to Galis et al., (2013). The lack of standards as well as the complexity of the SDN environment makes implementing DFR a challenge to some organisations.

Considering the newness of SDN and its complexity, other issues and challenges that may also directly or indirectly impact the DFR implementation in any SDN environment are:

- Interoperability,
- Reliability,
- Controller Placement (Controller Bottleneck) and
- Performance.

Note that these issues and challenges are also native to SDN but may in one way or the other have some impact on how DFR is implemented. The next section highlights some of the suggested countermeasures that organisations can embrace to enhance their ability to respond to cybersecurity incidents as well as help them in DFR implementation in SDN environments.

### **5. Suggested countermeasures to the issues and challenges impacting DFR implementation in SDN**

With reference to the issues and challenges identified and discussed in this study, this section offers insights into some of the potential countermeasures that organisations can embrace to enhance their ability to respond to the issues and challenges impacting DFR implementation in SDN environments. However, the authors also acknowledge at this point that the countermeasures discussed in this section are because of the purposeful sampled literature and not in any way a comprehensive list. Research still needs to be done to enhance or add to this list.



### **5.1 Develop DFR implementation policies**

Policies are mostly used to set the directional tone for different areas of a business organisation. However, because SDN is still considered new technology, policy development for DFR implementation needs careful planning as well as needs identification. Organisations need to gather as much information before any policy development process. The gathered information helps in drafting and reviewing the policy before the final implementation and subsequent reviews. A Well-defined SDN implementation policy can be a key asset to providing the basis for an organisation to analyse how to get from their existing legacy network systems to having a fully functional forensic-ready SDN environment.

### **5.2 Develop SDN standards**

In this context, the authors believe that, developing internationally accepted SDN standards can help organisations and other industry in applying world best practices. This, therefore, calls for collaboration between business organisations and different industry stakeholders to develop internationally accepted SDN standards that can ease the process of implementing DFR in SDN environments. Standards will also help organisations manage cybersecurity incidents with ease as well as maximize the potential use of digital evidence while minimizing digital forensic investigation costs in SDN environments.

### **5.3 Managing backward compatibility and avoiding lengthy downtime**

With the advancement in technology, backward compatibility allows for interoperability with older but existing legacy network systems. This is important because organisations can leverage the power of their new systems while still able to transact using old legacy network systems. For this reason, business organisations should consider, backward compatibility when considering a transition to help reduce the cost associated with network downtime, limit service disruption, and reduce possible network security risks (QuoteColo, 2016) especially during the process of implementing DFR. Besides backward compatibility, creating redundancy as well as preparing an organisation for disaster recovery ahead of time can help avoid lengthy downtimes as redundancy can help increase reliability as well as system performance.

### **5.4 Consider cost-benefit analysis for DFR implementation**

With cost-benefit analysis, organisations can estimate the strengths and weaknesses of alternative approaches to implementing DFR in SDN environments. This way organisations can determine what different options exist that can provide the best approach to achieving DFR implementation while managing their budget constraints as well as preserving savings.

### **5.5 Training of personnel**

Both old and new personnel should continually be trained as new technology emerge. Training helps organisation personnel comply with new forensic readiness best practices that help them address existing issues and challenges impacting the implementation of DFR in SDN. Training also ensures that personnel are aware of any existing or new standards, policies, and procedures to be used before, during, and after a digital investigation process. More countermeasures exist beyond what is discussed in this paper and every organisation should consider and explore other existing options including those not mentioned in this study. The next section concludes this paper.

## **6. Conclusion and future work**

In this paper, the authors have discussed different issues and challenges impacting the implementation of DFR in SDN environments. However, the paper has also highlighted the different countermeasures that organisations can embrace to enhance their ability to respond to cybersecurity incidents as well as help them during the transition process to SDN and more especially during DFR implementation in SDN environments. The presentation in this paper can, for example, help digital forensic practitioners, law enforcement agencies, as well as organisations in developing dynamic and proactive countermeasures to deal with the identified issues and challenges impacting DFR implementation in SDN. However, more research is still needed to provide directions on all the identified issues and challenges as well as the suggested countermeasures to the issues and challenges impacting DFR implementation in SDN. As part of future research, the authors plan to develop a DFR framework that can help ease the implementation of DFR in SDN as well as show how well organisations can deal with some of the identified issues and challenges in this study.

## Acknowledgements

The work has been supported by the Cyber Security Research Centre Limited whose activities are partially funded by the Australian Government's Cooperative Research Centres Programme

## References

- Ademu, I. O., Imafidon, C. O., & Preston, D. S. (2011). A new approach of the digital forensic model for digital forensic investigation. *Int. J. Adv. Comput. Sci. Appl*, 2(12), 175-178.
- Amin, R., Reisslein, M., and Shah, N. (2018). "Hybrid SDN Networks: A Survey of Existing Approaches," in *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3259-3306, Fourth quarter 2018, DOI: 10.1109/COMST.2018.2837161.
- Antonio P., and Labuschagne, L., (2012). A conceptual model for digital forensic readiness. *Proceedings of the ISSA Conference*; 2012Aug. 15-17, Johannesburg, SA. IEEE Publishers, 2012; pp.1-8
- Aric, T., (2020) SDN Network (Software Defined Network OpenFlow Protocol) Overview – The Ultimate Guide!! (What Is SDN And How It Really Works?). Available at: <https://electronicsguide4u.com/sdn-network-software-defined-network-openflow-protocol-what-is-sdn/> [Accessed on 29th January 2021]
- Desai, A.M, Fitzgerald, D., Hoanca, B., (2009). Offering a digital forensics course in Anchorage, Alaska. *Inform Syst Edu J* 2009;7(35); <http://isedj.org/7/35/>
- DFMAG (2020). Forensic Readiness. Available at: [https://digitalforensicsmagazine.com/index.php?option=com\\_content&view=article&id=916](https://digitalforensicsmagazine.com/index.php?option=com_content&view=article&id=916) [Accessed on 28th January 2021]
- Elyas, M., Ahmad, A., Maynard, S. B., & Lonie, A. (2015). Digital forensic readiness: Expert perspectives on a theoretical framework. *Computers & Security*, 52, 70-89.
- Galis A, Clayman S, Mamatas L, Rubio Loyola J, Manzalini A, Kuklinski S, Serrat J, Zahariadis T. Softwarization of Future Networks and Services - Programmable Enabled Networks as Next Generation Software Defined Networks. *IEEE SDN for Future Networks and Services (SDN4FNS)* 2013.
- ISO/IEC 27037:2012 — Information technology — Security techniques — Guidelines for identification, collection, acquisition, and preservation of digital evidence.
- Kandoi, R., and Antikainen, M. (2015). "Denial-of-service attacks in OpenFlow SDN networks," 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), Ottawa, ON, 2015, pp. 1322-1326, DOI: 10.1109/INM.2015.7140489.
- Karie, N. M., & Karume, S. M. (2017). Digital forensic readiness in organisations: Issues and challenges. *The Journal of Digital Forensics, Security and Law: JDFSL*, 12(4), 43-53.
- Karie, N. M., & Kebande, V. R. (2016). Building ontologies for digital forensic terminologies. *International Journal of Cyber-Security and Digital Forensics, (IJCSDF)* 5(2), pp.75-83.
- Kebande, V. R., Mudau, P. P., Ikuesan, R. A., Venter, H. S., & Choo, K. K. R. (2020). Holistic digital forensic readiness framework for IoT-enabled organisations. *Forensic Science International: Reports*, 2, 100117.
- Kemp, (2020). SDN Datapath. Available at: <https://kemptechnologies.com/au/glossary/sdn-datapath/> [Accessed on 29th January 2020]
- Khan, S., Gani, A., Wahab, A. W. A., Abdelaziz, A., Ko, K., Khan, M. K., & Guizani, M. (2016). Software-defined network forensics: Motivation, potential locations, requirements, and challenges. *IEEE Network*, 30(6), 6-13.
- Köhn, M., Olivier, M. S., & Eloff, J. H. (2006, July). Framework for a Digital Forensic Investigation. In *ISSA* (pp. 1-7).
- KPMG, (2015). Achieving Digital Forensic Readiness. Available at: <https://assets.kpmg/content/dam/kpmg/pdf/2016/03/Achieving-Digital-Forensic-Readiness-12-9-2015.pdf> [Accessed on 28th 01, 2021]
- Lagrasse, M., Singh, A., Munkhondya, H., Ikuesan, A., & Venter, H. (2020, March). Digital forensic readiness framework for software-defined networks using a trigger-based collection mechanism." In *Proceedings of the 15th International Conference on Cyber Warfare and Security, ICCWS* (pp. 296-305).
- Microsoft (2020). Network Policies. Available at: <https://docs.microsoft.com/en-us/windows-server/networking/technologies/nps/nps-overview> [Accessed on 16th February 2021]
- Mohay, G. (2005). Technical challenges and directions for digital forensics. In *First International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE'05)* (pp. 155-161). IEEE.
- Mugitama, S. A., Cahyani, N. D. W., and Sukamo, P. (2020). "An Evidence-Based Technical Process for OpenFlow-Based SDN Forensics," In the proceedings of the 8th International Conference on Information and Communication Technology (ICoICT), Yogyakarta, Indonesia, 2020, pp. 1-6, DOI: 10.1109/ICoICT49345.2020.9166215.
- Munkhondya, H., Ikuesan, A. R., & Venter, H. S. (2020). A case for a dynamic approach to digital forensic readiness in an SDN platform. In *International Conference on Cyber Warfare and Security* (pp. 584-XVIII). Academic Conferences International Limited.
- Munkhondya, H., Ikuesan, A., & Venter, H. (2019, February). Digital forensic readiness approach for potential evidence preservation in software-defined networks. In *ICCWS 2019 14th International Conference on Cyber Warfare and Security: ICCWS* (Vol. 268).
- ONF (2013). SDN Architecture Overview. Version 1.0. Available at: <https://opennetworking.org/wp-content/uploads/2013/02/SDN-architecture-overview-1.0.pdf> [Accessed on 29th January 2021]

- Park, S., Kim, Y., Park, G., Na, O., & Chang, H. (2018). Research on digital forensic readiness design in a cloud computing-based smart work environment. *Sustainability*, 10(4), 1203.
- QuoteColo (2016). Top Five Challenges Facing SDN. available at: <https://www.quotecolo.com/top-five-challenges-facing-sdn/> [Accessed on 16th February 2021]
- Reddy, K., Venter, H.S. & Olivier, M.S. Using time-driven activity-based costing to manage digital forensic readiness in large organisations. *Inf Syst Front* 14, 1061–1077 (2012). <https://doi.org/10.1007/s10796-011-9333-x>
- Rosencrance, L., English, J., and Burke, J. (2020). software-defined networking (SDN). Available at: <https://searchnetworking.techtarget.com/definition/software-defined-networking-SDN#> [Accessed on 29th January 2021]
- Rowlingson, R. (2004). A ten-step process for forensic readiness. *International Journal of Digital Evidence*, 2(3), 1-28.
- SDx, (2015). Understanding the SDN Architecture - SDN Control Plane & SDN Data Plane. Available at: <https://www.sdxcentral.com/networking/sdn/definitions/inside-sdn-architecture/#> [Accessed on 29th January 2021]
- Selamat, S. R., Yusof, R., & Sahib, S. (2008). Mapping process of digital forensic investigation framework. *International Journal of Computer Science and Network Security*, 8(10), 163-169.
- Sezer, S., Scott-Hayward, S., Chouhan, P. K., Fraser, B., Lake, D., Finnegan, J., ... & Rao, N. (2013). Are we ready for SDN? Implementation challenges for software-defined networks. *IEEE Communications Magazine*, 51(7), 36-43.
- Spiekermann, D., & Eggendorfer, T. (2016). Challenges of network forensic investigation in virtual networks. *Journal of Cyber Security and Mobility*, 15-46.
- Valjarevic, A., & Venter, H. S. (2012, August). Harmonised digital forensic investigation process model. In 2012 Information Security for South Africa (pp. 1-10). IEEE.
- Valjarevic, A., & Venter, H. S. (2015). A comprehensive and harmonized digital forensic investigation process model. *Journal of forensic sciences*, 60(6), 1467-1483.
- VMware (2021). Software-Defined Networking (SDN). Available at: <https://www.vmware.com/topics/glossary/content/software-defined-networking> [Accessed on 27th January 2021]

# Cyber Wargaming on the Strategic/Political Level: Exploring Cyber Warfare in a Matrix Wargame

Thorsten Kodalle

The Bundeswehr Command and Staff College, Hamburg, Germany

[thorstenkodalle@bundeswehr.org](mailto:thorstenkodalle@bundeswehr.org)

[thorstenkodalle@hotmail.com](mailto:thorstenkodalle@hotmail.com)

DOI: 10.34190/EWS.21.500

**Abstract:** NATO understands cyber within a cognitive, virtual and physical domain and on technical, tactical, operational, strategic and political levels. The NATO SAS 129 Research Task Group (RTG) "Gamification of Cyber Defence/Resilience" explores the advantages of gamification, especially Serious Games in the form of wargames, card-driven games and matrix wargames to support training and education in all domains and on all these levels. Within their task is the development of specific prototypes for these specific domains and levels. The Bundeswehr Command and Staff Course of the German Armed Forces implemented within their competence-based training and education system the Matrix Wargame "Kaliningrad 2018" (MWG Kaliningrad 2018) in 2018 for security policy education on the strategic and political level. MWG Kaliningrad 2018 was used several times in the Basic Staff Officer Course and the General Staff Officer Course National but also with students from the Hamburg University of Technology (TUH). This paper describes the history of the implementation of the game, evaluations from the course participants and insights from the most recent research and development approaches to develop a "Global Matrix Wargame" with a particular emphasis on information operations, information warfare and cyber warfare on the semantical level. It will examine how a matrix wargame can be a practical approach to reach specific cyber-related cognitive learning goals and appreciate a whole of government and whole of society approach to cyber resilience. It describes a good practice approach to a research and development and education (R&D&E) approach to discuss resilience in an open society. This is the second of three articles for the Conference Proceedings of ECCWS 2020. There are redundancies in the introduction, and the first article examines terminological uncertainties more.

**Keywords:** game-based learning (GBL), matrix wargames, cyber warfare on the semantic level, information operations

---

## 1. Introduction into NATO SAS 129: Background, objectives, topics, deliverables, administrative framework

The North Atlantic Treaty Organization System Analysis and Study (NATO SAS) 129 Research Task Group (RTG) "Gamification of Cyber Defence/Resilience" started its activity on 12 June 2017 and the current activity end date is 12 June 2021. A NATO SAS RTG is working within the framework of the NATO Science and Technology Organization (NATO STO), and their activity is proposed in a Technical Activity Proposal (TAP) (NATO Science and Technology Organization (STO) 2016) that needs to be approved. NATO SAS 129 TAP was approved in 2016 and described background and justification, objectives, topics, deliverables and the administrative framework of NATO SAS 129. This article will put a particular focus on the demonstrator that was developed at the Bundeswehr Command and Staff College: The Cyber Resilience Card Game (CRCG).

### 1.1 Background and justification (relevance to NATO)

The Wales Summit Declaration in 5 September 2014 (NATO 2014) provides a basis of justification for this research task group: "We are committed to developing further our national cyber defence capabilities, and we will enhance the cybersecurity of national networks upon which NATO depends for its core tasks, in order to help make the Alliance resilient and fully protected. Close bilateral and multinational cooperation plays a key role in enhancing the cyber defence capabilities of the Alliance. ... Technological innovations and expertise from the private sector are crucial to enable NATO and Allies to achieve the Enhanced Cyber Defence Policy's objectives. We will improve the level of NATO's cyber defence education, training, and exercise activities." (NATO 2014) Since the Wales Summit NATO issued a cyber pledge and designated Cyber as the 5<sup>th</sup> domain on the Warsaw Summit in 2016 (NATO 2016). For NATO a serious cyber-attack could trigger Article 5 (Stoltenberg 2019), and on the Brussels Summit, 2018 (NATO 2018a) NATO allies agreed to set up a new Cyberspace Operations Centre (NATO 2019).

In 2016 the TAP predicted that "Cyber threats and attacks will continue to increase in numbers, sophistication and the potential damage and this activity will contribute to cyber defence resilience in modern NATO environments." (NATO Science and Technology Organization (STO) 2016). This prediction turned out to be accurate; cyberspace is always on, and "NATO is a target three times over." (Omand 2019), not only the networks

of the organisation are targets but also NATO members and soldiers own mobile devices (Grove et al. 2017). Individual soldiers are particularly vulnerable targets (Bay and Biteniece 2019) and can be “catfished” (Lapowsky 2019). NATO SAS 129 developed this assessment further based on NATO expectations that the future theatre of war is expected to be mega-cities (Strategic Analysis Branch 2017, p. 38), where the environment is rich with cyber assets and is developing a Multi-Domain Future Urban Wargame addressing these issues on the tactical and operational level. However, simple cyber hygiene efforts are still the baseline of cyber defence and resilience according to NATO General Secretary Stoltenberg: “Some of the biggest cyber-attacks have only been possible because of human error. Such as picking up an infected USB Drive placed in a car park, and plugging it into a computer. Or clicking on a bad link in a ‘phishing’ email. It is time we all woke up to the potential dangers of cyber threats.” (NATO 2018c). The TAP recognised in 2016 that “many real-world solutions are available for training and education of cyber experts, there is a lack of training and education of cyber defence/resilience in general. Not many solutions are available for training and education of clients such as end-users, policymakers and military decision-makers.” Today there is a significant amount of open-source games available, that specifically target young end-users (Coenraad et al. 2020).

“Conventional methods for raising general awareness is often either costly or ineffective. Therefore one of the possible solutions for training and education is developing serious games and gamification applications.” This 2016 TAP statement still rings true today. Advertisement for professional Game-Based Learning (GBL)/Serious Games solutions are built on the premise that PowerPoint presentations for teaching, training, and raising awareness are boring and ineffective. Humans are still the weakest link in any cyber defence (Spatz 2017), and the low level the adoption of multi-factor authentication (MFA) or two-factor authentication (2FA) is still a concern (Das et al. 2019). Compared with the required vaccination rate for measles and pertussis (92-96%), rubella (84-88%) and mumps (88-92%) (Anderson and May 1985) we are far from herd immunity although the adoption rate of 2FA increased from 2017 (28%) significantly to 2019 (53%). Also, user awareness rose significantly from 44% to 77% (Engler 2019). However, opinions differ on the effectiveness (Colnago et al. 2018) (Covello 2019). There are also different national attitudes concerning providing 2FA for customers from the private sector. The willingness in Germany is lower (t3n Redaktion 2019) than in the U.S. (ThumbSignIn et al.). In the end, our society still provides a big attack surface.

It is assumed in the TAP that Serious Games and Gamification can contribute to solving this problem. It is also assumed “that games are available across platforms and can be designed in a way that it attracts the general audience”. For NATO as an organisation, it is essential to understand complex cyber resilience/defence/incident management scenarios. Based on the premise, “that gamification techniques can be useful in training and education regarding different cyber defence/resilience scenarios in a joint and high-pressure environment” the conclusion is drawn “thus, gamification provides opportunities to understand the possibilities inherent in cyber defence and train or educate people while they are having fun, contributing to the goals set forth by the Wales Summit.”

## **1.2 Objective(s)**

The stated objective of the TAP is “To effectively enhance information security and cyber defence education and training through the use of serious gaming and gamification approaches.”

## **1.3 Topic to be covered**

In the first 2016 TAP the proposed work consists of three main topics:

- 1) The definition of Serious Game and Gamification and the examination of advantages and disadvantages, common problems during development, gamification characteristics, game mechanics and technologies, and defence applications.
- 2) The understanding of the big picture of cyber defence and resilience as a baseline for the specification and prioritisation of cybersecurity subjects and user groups that can benefit from utilisation of Gamification and Serious Game applications, classification of operations and decisions in cyber defence and resilience, and examples of cybersecurity training and education.
- 3) The development of Gamification and Serious Game methodology guidelines for cyber defence and resilience, including the implementation of prototype demonstrations.

### 1.4 Deliverables

NATO SAS 129 is supposed to submit a final technical report documenting findings on gamification, describing cyber defence and resilience baseline information, and game methodology guidelines with prototypes developed.

### 1.5 Administrative framework: Lead nation, team leader and participants

The lead nation for NATO SAS 129 is Turkey, chaired by Mr Levent Berke Capli. Nations and organisations participating are USA, Great Britain, Germany, Netherlands, Turkey, and NATO Cooperative Cyber Defence Center of Excellence (NATO CCD COE) whose efforts include enhancing information security and cyber defence education awareness and training.

## 2. Terminology: About cyber and games

This paper offers only an overview of the current discussion of the terminology used in this paper. There is a terminological grey area between *Wargaming*, *Serious Games*, *Game-Based Learning*, and depending on the usage also *Gamification*. For a more comprehensive discussion of the terms Wargaming, Serious Games, Game-Based Learning and Gamification read the accompanying article “Cyber Wargaming on the technical/tactical level: The Cyber Resilience Card Game (CRCG)” in the conference proceedings of ECCWS 2020.

Picture Strategic Thinking

### 2.1 Cyber defence and cyber resilience

There is still an absence of an internationally accepted and harmonised definition of cybersecurity and definitions of the related concepts (Buřita 2019). The author, as a certified NATO Cyber Defense Advisor offers his understanding of Cyber as a domain with different subdomains (*Cyber Domains*) and different levels (*Cyber Levels*) (see Figure 1).

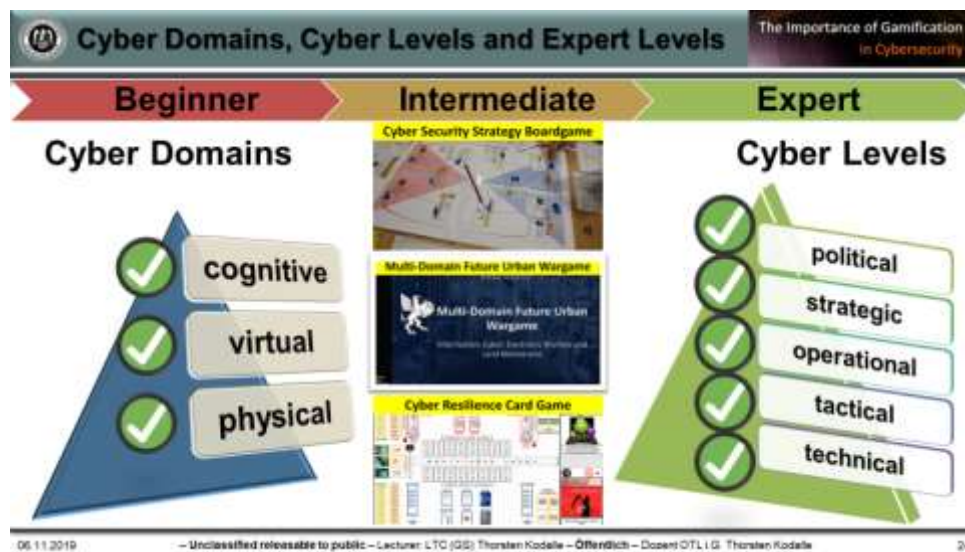


Figure 1: Cyber domains, cyber levels and expert levels

Figure 1 is a crucial result of the research and development (R&D) Workshop NATO SAS 129 conducted in a joint endeavour with the German Institute for Defence and Strategic Studies (GIDS) 18 -20 June 2019. Three prototypes developed within the framework of NATO SAS 12) are featured in the middle of the slide with the CRGC at the bottom. Three categories *Beginner*, *Intermediate* and *Expert* at the top refer to the level of expertise a participant in these games has, either as a cyber expert and as a gamer.

On the left, there are three *Cyber Domains*. The *physical* domain includes actual hardware like computers, routers, wires et al. The "virtual" domain includes intangible software and in example logical layers, protocols et al. The *cognitive* domain includes all human mental or mind-based, conscious and unconscious junctures to the other domains. Specific attack vectors focus on specific domains. Social engineering is implemented in the cognitive domain. Logical bombs are planted in the virtual domain and centrifuges in uranium enrichment

facilities are blown up on the physical level. These cyber domains also differentiate into the *technical, tactical, operational, strategic* and *political Cyber Level* (at the right of Figure 1). The private mobile device of a NATO soldier is hardware (*physical domain*) on the tactical level. A command and control (C<sup>2</sup>) server in a corps headquarter (HQ) could be hardware (*physical domain*) or virtualised in the cloud (*virtual domain*) on the operational level. The email system of the government is in the virtual domain and on the political level. Disinformation campaigns are effective in the cognitive domain on the tactical level, maybe with huge political effects. Thinking about cyberspace always requires a holistic approach, because everything is connected, literally. The different stages of a unified kill chain (UKC) illustrate this interconnectedness: 1. Reconnaissance, 2. Weaponisation, 3. Delivery, 4. Social Engineering, 5. Exploitation, 6. Persistence, 7. Defence Evasion, 8. Command & Control, 9. Pivoting, 10. Discovery, 11. Privilege Escalation, 12. Execution, 13. Credential Access, 14. Lateral Movement, 15. Collection, 16. Exfiltration, 17. Target Manipulation, 18. Objectives (Pols 2017, p. 87). Therefore the concept of cyber resilience (or resilience in general) is more important. Resilience in this context means “Having sufficient capability, capacity, and will to endure adversity over time, retain the ability to respond, and to recover quickly from strategic shocks or operational setbacks. (NATO 2018b, A62). Resilience should include the idea for an organisation to be still functional, even suffering successful attacks, which close to “the ability of a functional unit to continue to perform a required function in the presence of faults or errors. NATO Adopted 2005-03-01 (NATO Standardization Office (NSO) 12/3/2019, p. 3731).

### **3. History of development and implementation**

#### **3.1 First idea**

A Matrix Wargame is a particular variation of manual wargame with a special emphasis on a structured discussion of player moves. It is more like a pen and paper roleplay and has no relation to any mathematical ideas discussed in game theory. Chris Engle invented Matrix Games (Engle 2019, 2018). John Curry provided several books around the topic of Matrix Wargaming (Curry and Perla 2014; Curry and Price 2017; Curry et al. 2018; Curry and Price 2013). The author was introduced to Matrix Wargames end of 2017 by a delegation of the U.S. Army War College during their visit at the Bundeswehr Command and Staff College. The method was embraced by some lectures, including the author and implemented into seminar-style matrix wargaming. A colleague of the author developed his Matrix Wargame to cover United Nations missions, particularly in Africa and played it with international students from Africa. The author started with the provided “Kaliningrad 2018” (Brynen 2016b, 2016a, 2018) Matrix Wargame, which students adapted throughout several classes. It is particularly useful to teach the DIME/PMESII framework (see below) for security policy with a particular focus on the information domain, covering topics from critical infrastructure and cyber warfare on the semantic level. This approach covers the physical and virtual cyber domain (like attacks on industrial control systems (ICS) to disrupt society on a vast scale) and the cognitive domain (like disinformation and influence campaigns for hacking elections). These topics are covered on the strategic and political level.

##### *3.1.1 Covering security (cyber) policy with an MWG in a holistic framework: DIME And PMESII*

The DIME/PMESII paradigm is the applied framework (Hartley III 2017, pp. 99–106). Although “The origins of the DIME/PMESII paradigm are unclear” (Hartley III 2017, p. 99) and acknowledging its limitations, the DIME/PMESII model is useful, reflected across many military doctrines. “Essentially, all models are wrong, but some are useful” (Box 1976). There are alternatives to DIME and PMESII. Some argue that the linear structure of the PMESII structure only reveals the “what” and not the “why” of complex systems (Hartley III 2017, p. 106). An alternative would be MIDLIFE (Military, Informational/Intelligence, Diplomatic, Legal (Law Enforcement), Infrastructure, Finance, Economics) as a concept of national power. MIDLIFE was presented in the expired United States Army doctrine (FMI 3-07.22) for Counterinsurgency Operations, October 2004 (US Army Field Manual Interim 2004). Nevertheless, DIME is still very prominent in U.S. Doctrine – and they put a particular focus on the “I” (Bishop 2018a). Also, the new NATO military strategy uses DIME as the concept for national powers (McLeary 2019). PESTEL (the separation of the political, economic, social, technological, environmental and legal domains) is used as an analytical tool in combination with the SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis (Mullerbeck 2015). According to Konstantin Khomko, combining these schemas (DIME, MIDLIFE and PESTEL) would offer the most balanced approach to articulating national power elements, especially regarding contemporary issues (Khomko 2019). Due to the prominence of DIME in the public debate (Bishop 2018b), the DIME framework is still a useful framework for conceptualising security policy and feature at the centre of a Matrix Wargame on the political level.

3.1.2 The DIME concept

The DIME concept groups the many instruments of power a nation-state can muster into four easy to remember categories as follows:

CATEGORY	Characteristics	Actions
Diplomacy	Soft power	e.g. negotiations and agreements
Information	Soft power	e.g. gaining and controlling information
Military	Hard power	e.g. usage of security forces including police
Economics	Hard power	e.g. sanctions and trade war

Figure 2: DIME Categories. characteristics and actions adopted after (Hartley III 2017, p. 103)

Two categories collect the soft power instruments or levers of power: Diplomacy and Information. Diplomatic power rests on negotiations and agreements. Information power lies in gaining information from others and in controlling the information desired by others. Two categories collect the hard power instruments or levers of power: Military and Economics. Military power is a distinct component. The police and other instruments/actors from the executional power of a nation-state are also included in this category. Economic power is also a prominent component. The idea of framing military and economic power as *hard power* instruments and diplomatic power and information power as *soft power*, however, narrows the view of the application of these instruments. "I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail" (Maslow 1977). Maslow’s “law of the hammer” allows us to identify a cognitive bias, seeing military and economic levers as somehow ‘harder’ than diplomacy or information (Maslow 1977).

3.1.3 The PMESII concept

The six domains of PMESII (Political, Military, Economic, Social, Information and Infrastructure) can be understood as part of an operational environment (Hartley III 2017, p. 101). Each domain also further subdivides into a series of components, as follows:

DOMAIN	Sub-components
Political	Governance, the rule of law
Military	conflict, government, security (including Intelligence services)
Economic	agriculture, crime, energy, finance, governmental economic actions, employment
Social	basic needs, education, health, movement, safety
Informational	general information items, media, opinions, information operations
Infrastructure	business infrastructure, social infrastructure, energy infrastructure, government infrastructure, transportation infrastructure, water infrastructure

Figure 3: Key elements of the PMESII space after (Hartley 2017, pp. 99–103)

The PMESII concept combines the domain interactions into a system and creates a framework for operational design and joint planning. The planner has a frame of reference for collaboration with inter-organizational and multinational partners to determine and coordinate actions, fostering a comprehensive approach and providing a steppingstone for a Whole of Government-(WoG) approach. PMESII supports the identification of Centre of Gravity (CoG) on different levels, operational CoG and strategic CoG (Hartley III 2017, p. 101).

3.2 Implementation

The Matrix Wargame “Kaliningrad 2018” was implemented in the first class of the Basic Officer Staff Course (BLS) of 2018 (BLS 1-2018). The very first run followed the provided source material very carefully. However, the U. S. Army War College provided a plan for a one-day event with around ten expected moves by experts in the rank of full colonels. The target group of students in the BLS is on average in the rank of captain and has around 20 years less professional experience each which amounts in a class of 17 students (with 12 years average of professional experience equals 214 years total) compared to ten full colonels (with a combined professional experience of around 300 years) a deficit of almost 100 years professional experience. Therefore the goal of playing a Matrix Game with captains should be different than playing a Matrix Wargame with full colonels. Speed in execution was slowed down considerably and time for proper documentation in PowerPoint was added. In the first run, the Matrix Wargame took several days, and only five turns were made. By coincidence, a subject matter expert (former member of the MoD of Lithuania) was contributing voluntarily and advised the actor “The Baltics States, Poland and Sweden” on-premise. Another SME (Turkey) was supporting virtual within a WhatsApp group.



### **3.3 Iterations**

The BLS 2-2018 and BLS 1-2019 repeated the Matrix Wargame, directly followed by the then younger class of the General/Admiral National Staff Course (LGAN 2018 – the LGAN is a two-year course, and one class is always the older one, and the other is the younger class). In BLS 3-2019 and BLS 1-2020 “Kalinigrad 2020” was also used and further developed. One of the main milestones was the implementation of a Slack Workspace in February 2019. The LGAN played it first. Since then all Matrix Wargames are facilitated within the same Slack Workspace, and all player moves since then are saved and documented in the Slack Workspace which is a vast improvement compared to several WhatsApp groups for each actor. The author presented the advantages of facilitating a Matrix Wargame in a Slack Workspace at 2019 Connections UK and the 2019 European Conference on Game-Based Learning (ECGBL) (Kodalle 2019).

### **3.4 Status quo**

The last iteration in January 2020 took a particular focus on cyber warfare on the semantic level and attack vectors targeting the cognitive domain. The MWG was a preparation for the visit of the Bundeswehr Cyber Innovation Hub with a presentation and discussion of how to hack a nuclear powerplant and attacks on critical infrastructure via ICS and sensors. Students also visited the Ministry of Interior and the Ministry of Defence and discussed terrorism use of cyberspace, in particular, the weaponisation of social media (Kandemir et al. 2017). After this visit students gave presentations on their specific actors and specific cyber-related phenomenons, including deep fakes in relationship with artificial intelligence (AI). The students also created *learningsnacks* (at [www.learningsnacks.de](http://www.learningsnacks.de)) for distributing their knowledge in the form of microlearning content that is accessible through mobile devices.

## **4. Evaluation**

Students wrote on four occasions, After-Action Reports (AAR) by the official format of MoD reports. The critical question was, how students evaluated the Matrix Wargame as a tool for competency-based training? Competence-based training includes a self-reflection of the training as the last step in the comprehensive action a student takes (Simberg 2017).

### **4.1 Player evaluation**

The general player evaluation was very positive. Players were biased because they addressed their lecturer and tutor with their report, and they were up for evaluation. However, even with a strong incentive to provide positive feedback, students provided nuanced and constructive feedback and did not shy to criticise elements of gameplay, content and procedures of the MWG. They identified the trade-offs concerning jumping around the globe and not sticking to the provided map (NATO Eastern Flank featuring Kaliningrad, Nord Stream 2 and the Suwalki Gap (Brynen 2018)). They also requested more time for evaluation of other actors moves.

### **4.2 Facilitator evaluation**

The facilitator provided the student with a purpose to compensate for their strong bias for the positive feedback. The MWG is an educational, research and development project, where students can contribute in a meaningful way if they would provide constructive feedback. This provided the higher purpose of working on the system, and not only in the system. Students were tasked not to shy away from the critic, but to provide constructively. The facilitator observed highly motivated and deeply in the learning process involved students. All critic was constructive and helpful and provided significant input and recommendations for improving gameplay and game mechanics.

### **4.3 How to implement cyber topics in a matrix wargame**

The MWG provided a framework (DIME/PMESII) to systemise player moves and contextualised their actions. Within the provided framework, any move in any DIME domain can be analysed in the physical, virtual or cognitive cyber domain. Secret moves provided a space to play out hybrid warfare that created effects in the game only if they influence the gameplay at all. In the end, players revealed their secret moves sometimes with a great surprise for other actor groups.

#### 4.4 Cyber warfare on the semantic level

The author evaluated each move for effect on world opinion and internal cohesion of an actor based on the seminar discussion with the students. Moreover, players used secret moves to conduct hybrid warfare and tried to influence the population by framing the narrative or signalling cooperative behaviour while behaving confrontational in secret moves. Some moves explicitly targeted specific groups of the population and used social media. A typical way for resolving effects in an MWG is a trough of two six-sided dices (2 D6). Based on probability and provided rationals solid argumentation with accurate facts will provide a high probability off an effect taking place or the strength of an effect.

#### 4.5 Critical infrastructure

Critical infrastructure was targeted on all level, starting on the tactical level by sending the cleaning lady from a disenfranchised ethnic minority group with a USB stick to plant a logic bomb in the computer system of an LPG harbour facility to use it just in case. Actors attacked the financial sector of a competitor by targeting banks as a way to help their banking facility to grow in the local market. Player also used fake news to discredit political leaders and target the cohesion of the population of other actors.

### 5. Conclusion

A Matrix Wargame is a flexible and adaptable method to address several needs from analysis to education. The Matrix Wargame in this article was implemented for educational purposes and to raise awareness. In principle, any topic can be focused within the provided DIME/PMESII framework to contextualise nation-state actions within a comprehensive or integrated or WoG approach. Any wargame needs to have boundaries. Usually, there are boundaries in space and time, and the facilitator needs to analyse, what the most appropriated time frame and regional focus are to achieve the desired outcome. In this case, there is the exception of no particular regional focus, although the Kaliningrad region and the player sheets provide regional information and specific strategic goals. However, due to the demand of most recent security policy developments, and the intention to provide a high degree of freedom in their action, the players, went “out of area” very soon. The set boundary was a thematic one. Within the DIME/PMESII framework, the player focused on cyber-related actions and conducted hybrid moves, cyberattacks on critical infrastructure, disinformation campaigns and influence operations, bouncing all over the globe from Venezuela to the Street of Hormuz back to Kaliningrad. There is a trade-off between in-depth analysis and expert discussion and raising awareness and exploring the big picture of cyber on a global stage. Students enjoyed their freedom of intellectual movement, embraced the different actor perspectives and judged this method to be very motivating for educational purposes. For future development, the author will use design thinking to develop and provided digital learning materials for a holistic learning environment. Self-determination theory is the theoretical backbone for a growth mindset, that embraces intrinsic motivation and meaningful choices by players.

### References

- Anderson, Roy M.; May, Robert M. (1985): Vaccination and herd immunity to infectious diseases. In *Nature* 318. Available online at <https://www.nature.com/articles/318323a0.pdf>, checked on 4/1/2020.
- Bay, Sebastian; Biteniec, Nora (2019): The current digital arena and its risk to serving military personnel. In *NATO STRATCOM COE*, pp. 7–18, checked on 1/4/2020.
- Bishop, Donald M. (2018a): DIME, not DiME: Time to Align the Instruments of U.S. Informational Power. *The Strategy Bridge*. Available online at <https://thestrategybridge.org/the-bridge/2018/6/20/dime-not-dime-time-to-align-the-instruments-of-us-informational-power>, checked on 7/20/2019.
- Bishop, Donald M. (2018b): DIME, not DiME: Time to Align the Instruments of U.S. Informational Power. *The Strategy Bridge*. Available online at <https://thestrategybridge.org/the-bridge/2018/6/20/dime-not-dime-time-to-align-the-instruments-of-us-informational-power>, updated on 6/20/2018, checked on 10/2/2020.
- Box, George E. P. (1976): Science and Statistics. In *Journal of the American Statistical Association* 71 (356), pp. 791–799. DOI: 10.1080/01621459.1976.10480949.
- Brynen, Rex (2016a): Kaliningrad 2017 matrix game at the US Army War College. US Army War College. Available online at <https://paxsims.wordpress.com/2016/08/15/kaliningrad-2017-matrix-game-at-the-us-army-war-college/>, checked on 5/30/2019.
- Brynen, Rex (2016b): Kaliningrad 2017 playtest at NDU. US Army War College. Available online at <https://paxsims.wordpress.com/2016/06/07/kaliningrad-2017-playtest-at-ndu/>, checked on 5/30/2019.
- Brynen, Rex (2018): Matrix games for student learning at the US Army War College. US Army War College. Available online at <https://paxsims.wordpress.com/2018/05/31/matrix-games-for-student-learning-at-the-us-army-war-college/>, checked on 5/30/2019.

- Buřita, Ladislav (2019): Online glossary of cyber security. In Tiago Cruz, Paulo Simoes (Eds.): ECCWS 2019 - PROCEEDINGS OF THE 18TH EUROPEAN CONFERENCE ON CYBER WARFARE AND, vol. 2019-July. [S.l.]: ACPIL, pp. 72–77, checked on 4/1/2020.
- Coenraad, Merijke; Pellicone, Anthony; Ketelhut, Diane Jass; Cukier, Michel; Plane, Jan; Weintrop, David (2020): Learning Cybersecurity One Game at a Time: A Systematic Review of Cybersecurity Digital Games. under review.
- Colnago, Jessica; Devlin, Summer; Oates, Maggie; Swoopes, Chelse; Bauer, Lujo; Cranor, Lorrie; Christin, Nicolas (2018): "It's not actually that horrible". In CHI (Ed.): CHI 2018. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, April 21-26, 2018, Montreal, QC, Canada. the 2018 CHI Conference. Montreal QC, Canada, 21.04.2018 - 26.04.2018. CHI; Conference on Human Factors in Computing Systems. New York, NY: ACM, pp. 1–11. Available online at <https://www.archive.ece.cmu.edu/~lbauer/papers/2018/chi2018-2fa.pdf>, checked on 4/1/2020.
- Covello, Bob (2019): Opinion: Back to the Start for 2FA Adoption? Available online at <https://www.tripwire.com/state-of-security/security-awareness/back-to-the-start-for-2fa-adoption/>, checked on 4/1/2020.
- Curry, John; Engle, Chris; Perla, Peter P. (Eds.) (2018): The Matrix Games Handbook. Professional applications from education to analysis and wargaming. [Bristol]: The History of Wargaming Project (History of Wargaming Project).
- Curry, John; Perla, Tim Price (2014): Matrix games for modern wargaming. Developments in professional and educational wargames, innovations in wargaming, volume 2. San Bernardino, California: History of Wargaming Project (History of Wargaming Project).
- Curry, John; Price, Tim (2013): Dark Guest Training Games for Cyber Warfare Volume 1: Wargaming Internet Based Attacks (English Edition) Kindle Edition.
- Curry, John; Price, Tim (2017): Modern crises scenarios for matrix wargames. Kindle: History of Wargaming Project (History of Wargaming Project).
- Das, Sanchari; Wang, Bingxing; Tingle, Zachary; Camp, L. Jean (2019): Evaluating User Perception of Multi-Factor Authentication: A Systematic Review. Available online at <https://arxiv.org/pdf/1908.05901>.
- Engle, Chris (2018): WSS16 - Matrix Gaming. <https://www.facebook.com/thehistorynetwork>; The History Network. Podcasts, Wargames, Soldiers & Strategy. Available online at <http://thehistorynetwork.org/wss16-matrix-gaming/>, checked on 5/30/2019.
- Engle, Chris (2019): Free Engle Matrix Games. Available online at <https://sites.google.com/view/free-engle-matrix-games/home>, checked on 5/30/2019.
- Engler, Maggie (2019): State of the Auth. Available online at <https://duo.com/assets/ebooks/state-of-the-auth-2019.pdf>, checked on 4/1/2020.
- Grove, Thomas; Barnes, Julian E.; Hinshaw, Drew (2017): Russia Targets NATO Soldier Smartphones, Western Officials Say. In Wall Street Journal, 10/4/2017. Available online at <https://www.wsj.com/articles/russia-targets-soldier-smartphones-western-officials-say-1507109402>, checked on 4/1/2020.
- Hartley, Dean S. (2017): Unconventional conflict. A modeling perspective. Cham, Switzerland: Springer (Understanding complex systems).
- Hartley III, Dean S. (2017): Unconventional Conflict. A Modeling Perspective. Cham, s.l.: Springer International Publishing (Understanding complex systems).
- Kandemir, Berfin; Brand, Alexander; Heap, Ben; Allan, Iona; Ropsa, Inga (2017): Social Media in Operations. - a Counter-Terrorism Perspective, p. 28, checked on 1/13/2020.
- Khomko, Konstantin (2019): A nation needs more than a DIME. The Sir Richards Williams Foundation. THE CENTRAL BLUE. Available online at <http://centralblue.williamsfoundation.org.au/a-nation-needs-more-than-a-dime-konstantin-khomko/>, checked on 7/20/2019.
- Kodalle, Thorsten (2019): Hosting a Matrix Wargame in a Slack Workspace. Reading : 405-413, XVIII. Reading: Academic Conferences International Limited. (Oct 2019). In Lars Elbaek, Gunver Majgaard, Andrea Valente, Md. Saifuddin Khalid (Eds.): European Conference on Games Based Learning, XVIII. European Conference on Games Based Learning. University of Southern Denmark, 3 – 4 October 2019. Academic Conferences International Limited. XIII: Academic Conferences International Limited, pp. 405–413, checked on 2/10/2020.
- Lapowsky, Issie (2019): NATO Group Catfished Soldiers to Prove a Point About Privacy | . WIRED. Available online at <https://www.wired.com/story/nato-stratcom-catfished-soldiers-social-media/>, checked on 1/4/2020.
- Maslow, Abraham H. (1977): Die Psychologie der Wissenschaft. Neue Wege d. Wahrnehmung u.d. Denkens. München: Goldmann (Goldmann-Sachbücher, 11131).
- McLeary, Paul (2019): Dunford: Leaders Mull First NATO Strategy In Decades. Available online at <https://breakingdefense.com/2019/05/dunford-first-nato-strategy-okd-in-decades/>, checked on 7/20/2019.
- Mullerbeck, Eric (2015): SWOT AND PESTEL. Understanding your external and internal context for better planning and decision-making. UNICEF. Learning and Knowledge Exchange. Available online at [https://www.unicef.org/knowledge-exchange/index\\_83128.html](https://www.unicef.org/knowledge-exchange/index_83128.html), updated on 9/14/2015, checked on 7/20/2019.
- NATO (2014): Wales Summit Declaration. issued by the Heads of State and Government participating in the meeting of the North Atlantic Council in Wales. NATO. NATO.int. Available online at [https://www.nato.int/cps/en/natohq/official\\_texts\\_112964.htm?selectedLocale=en](https://www.nato.int/cps/en/natohq/official_texts_112964.htm?selectedLocale=en), updated on Last updated: 8/30/2018, checked on 11/10/2019.
- NATO (2016): Warsaw Summit Communiqué. Issued by the Heads of State and Government participating in the meeting of the North Atlantic Council in Warsaw, 8-9 July 2016. NATO. NATO.int. Available online at

## Thorsten Kodalle

- [https://www.nato.int/cps/en/natohq/official\\_texts\\_133169.htm?selectedLocale=en](https://www.nato.int/cps/en/natohq/official_texts_133169.htm?selectedLocale=en), updated on Last updated: 3/29/2017, checked on 11/10/2019.
- NATO (2018a): Brussels Summit Declaration. issued by the Heads of State and Government participating in the meeting of the North Atlantic Council in Brussels, 11-12 July 2018. NATO. NATO.int. Available online at [https://www.nato.int/cps/en/natohq/official\\_texts\\_156624.htm?selectedLocale=en](https://www.nato.int/cps/en/natohq/official_texts_156624.htm?selectedLocale=en), updated on Last updated: 8/30/2018, checked on 11/10/2019.
- NATO (2018b): FRAMEWORK FOR FUTURE ALLIANCE OPERATIONS. Available online at [https://www.act.nato.int/images/stories/media/doclibrary/180514\\_ffao18.pdf](https://www.act.nato.int/images/stories/media/doclibrary/180514_ffao18.pdf), checked on 1/4/2020.
- NATO (2018c): NATO - Opinion: Speech by NATO Secretary General Jens Stoltenberg at the Cyber Defence Pledge Conference (Ecole militaire, Paris), 15-May.-2018. NATO.int. Available online at [https://www.nato.int/cps/en/natohq/opinions\\_154462.htm](https://www.nato.int/cps/en/natohq/opinions_154462.htm), checked on 4/1/2020.
- NATO (2019): Cyber defence. NATO. NATO.int. Available online at [https://www.nato.int/cps/en/natohq/topics\\_78170.htm](https://www.nato.int/cps/en/natohq/topics_78170.htm), updated on Last updated: 9/6/2019, checked on 11/10/2019.
- NATO Science and Technology Organization (STO) (2016): Technical Activity Proposal. TAP, revised 2019. Source: NATO STO, checked on 11/10/2019.
- NATO Standardization Office (NSO) (12/3/2019): Terminology extracted from NATOTerm. Available online at <https://nso.nato.int/natoterm/Web.mvc>, checked on 1/15/2020.
- Omand, David (2019): Cyber Threats to NATO. In Ares & Athena (16). Available online at <https://chacr.org.uk/docs/20191121-Ares-and-Athena-16.pdf>, checked on 4/1/2020.
- Pols, Paul (2017): The Unified Kill Chain: Modeling Fancy Bear Attacks. Designing a Unified Kill Chain for analyzing, comparing and defending against cyber attacks, 2595 AN The Hague. Cyber Security Academy (CSA). Available online at [https://www.csacademy.nl/images/scrypties/2018/Paul\\_Pols\\_-\\_The\\_Unified\\_Kill\\_Chain\\_1.pdf](https://www.csacademy.nl/images/scrypties/2018/Paul_Pols_-_The_Unified_Kill_Chain_1.pdf), checked on 11/10/2019.
- Simberg, Martin (2017): Kompetenzorientierte Ausbildung in der Bundeswehr. Available online at <https://www.fueakbw.de/index.php/de/aktuelles/218-kompetenzorientierte-ausbildung-in-der-bundeswehr-testlauf-im-basislehrgang-stabsoffizier>, checked on 5/30/2019.
- Spatz, Scott (2017): How To Add Security To Your Offering. Available online at <https://www.mspinsights.com/doc/how-to-add-security-to-your-offering-0003>, checked on 8/1/2020.
- Stoltenberg, Jens (2019): NATO WILL DEFEND ITSELF. The alliance will guard its cyber domain—and invoke collective defence if required. In Prospect October 2019, p. 4. Available online at [https://www.prospectmagazine.co.uk/content/uploads/2019/08/Cyber\\_Resilience\\_October2019.pdf](https://www.prospectmagazine.co.uk/content/uploads/2019/08/Cyber_Resilience_October2019.pdf), checked on 11/10/2019.
- Strategic Analysis Branch (2017): Strategic Foresight Analysis. 2017 Report. NATO HQ SACT Strategic Plans and Policy. Available online at [https://www.act.nato.int/images/stories/media/doclibrary/171004\\_sfa\\_2017\\_report\\_hr.pdf](https://www.act.nato.int/images/stories/media/doclibrary/171004_sfa_2017_report_hr.pdf), checked on 1/17/2020.
- t3n Redaktion (2019): Deutsche Firmen drücken sich vor Zwei-Faktor-Authentifizierung. Available online at <https://t3n.de/news/deutsche-firmen-druecken-1203083/>, updated on 1/4/2020, checked on 4/1/2020.
- ThumbSignIn; One World Identity; Gluu: Customer Authentication Practices 2019 Survey. Available online at <https://thumbsignin.com/customer-authentication-report-2019/>, checked on 4/1/2020.
- US Army Field Manual Interim (2004): FMI 3-07.22, Counterinsurgency Operations. Available online at <https://fas.org/irp/doddir/army/fmi3-07-22.pdf>, checked on 7/20/2019.

# Cyber-Threat Analysis in the Remote Pilotage System

Tiina Kovanen, Jouni Pöyhönen and Martti Lehto

University of Jyväskylä, Finland

[tiina.r.j.kovanen@jyu.fi](mailto:tiina.r.j.kovanen@jyu.fi)

[jouni.a.poyhonen@jyu.fi](mailto:jouni.a.poyhonen@jyu.fi)

[martti.j.lehto@jyu.fi](mailto:martti.j.lehto@jyu.fi)

DOI: 10.347190/EWS.21.067

**Abstract:** Fairway pilotage is advancing toward a more digitalized future where an automated remote pilotage system such as ePilotage (a remote piloting system of systems) is possible. ePilotage is an example of a system in which an increased number of digital solutions are entering new environments where traditional engineering solutions are still in use. This development introduces increased risk of a malicious cyber adversary taking deliberate actions against the system. Cyber threats are a multidimensional phenomenon with many aspects to consider for technical and non-technical audiences. Often, the threat is perceived from a limited number of viewpoints, and important discoveries may be missed. Many organizations adapt public models to their needs and in the process lose some interoperability between different organizations. This is a compromise that has to be weighed. ePilotage is a very special viewpoint as it incorporates a large network of separate systems and stakeholders. By examining the impacts of cyber-threat actions in this connected environment, we found that the impacts affecting one subsystem are propagated to affect other systems. This effect combined with time-criticality implies that cooperation among subsystem stakeholders is essential. This requires common situational awareness to support fast and precise reactions. For this, there must be a common language to describe the situation. This is achieved by creating and using a common methodology for cyber-threat analysis. At the strategic level, there must be a description of the situation with non-technical terminology. This includes, for example, discussion of the attacker's motivations by studying different types of adversaries, such as cyber vandalism and cybercrime. At the tactical level, there must be more technical information on the threat actor's tactics, techniques, and procedures. By examining the cyber-threat actors' features and by combining them to known cyber-attack tactics, techniques and procedures, scenarios for ePilotage can be created. These scenarios provide information on the possible threats concerning this specific environment and its protection.

**Keywords:** maritime autonomy solution, ePilotage, cyber security, cyber threat analysis, scenarios

---

## 1. Introduction

Growing interdependencies across critical infrastructure systems, specifically, reliance on information and communications technologies, have increased potential vulnerabilities to cyber threats and potential consequences resulting from the compromise of underlying systems or networks (DHS, 2013). In the increasingly interconnected fairway pilotage system of systems (SoS), the potential impacts increase with these interdependencies and the ability of a diverse set of threats to exploit them.

An automated remote pilotage system, ePilotage (DIMECC, 2020), contains fairway and ship systems and control centers as the main functionalities. Vessel Traffic Service (VTS), other vessels' systems, weather forecasters, and stakeholder operators act as supporting functionalities (Pöyhönen, Kovanen & Lehto, 2021). Traditional marine traffic is very dependent on insecure communication, such as VHF and unencrypted emails. Vessel information is available online, including the identity, type of cargo, location, route, and schedule. Multiple attack demonstrations have been published for maritime traffic, including technological vulnerabilities in Automatic Identification System (AIS) and possibilities for disturbing Global Navigation Satellite System (GNSS) (Kovanen, Pöyhönen & Lehto, 2021). If a new ePilotage is built on top of this insecure foundation, possibilities for attacks will remain.

Impact scenarios for cyber threats in ePilotage have been investigated (Kovanen et al., 2021), and several key aspects specific to ePilotage were discovered. The effects of the impacts are not contained within one subsystem. The SoS includes information and communication technology (ICT) and industrial control system (ICS) components. Time-criticality in detection and countermeasures were introduced from the narrow safe passage the in fairway environment. Systems are separated in multiple physical locations, and cyber-physical effects can manifest. These findings provide the seed for evaluation and development of a cyber-threat analysis suitable for ePilotage. However, this impact analysis aids only in very high abstraction level planning as it does not present technical details. Therefore, this paper goes a step further in scenario creation utilizing a more detailed threat model.

The remainder of the paper is organized as follows: The second section proposes constructs for a cyber-threat model for ePilotage needs. The third section presents a rationalized method for creating scenarios for ePilotage and describes two scenario examples. The fourth section analyzes the created scenarios. The fifth section includes the conclusions and suggestions for future work.

## **2. Cyber-threat model for ePilotage**

Threats in cyberspace are difficult to define as it is hard to identify the source of attacks and the motives that drive them, or even to foresee the course of an attack as it unfolds. Identifying cyber threats is further complicated by the difficulty in defining the boundaries between national, international, public, and private interests. Because threats in cyberspace are global and involve rapid technological developments, the struggle to meet them is ever-changing and increasingly complicated (Lehto, 2013).

The word *threat* is used to refer to the adversary or the attack depending on the context (Bodeau, McCollum & Fox, 2018). The definition and usage of the term “threat modelling” are ambiguous (Xiong & Lagerström, 2019) and can be employed in multiple abstraction levels (Bodeau & McCollum, 2018). Moreover, a comprehensive ontology for cyber-threat intelligence that incorporates all relevant data and abstraction levels is lacking (Mavroeidis & Bromander, 2018). This situation makes it difficult to communicate with other organizations or within organization between different abstraction levels used by different roles, such as strategic planner and day-to-day incident responder.

Many of the current models view cyber threats as a technological problem in system or software development. For example, the most widely used threat model, STRIDE, bases their threat categories on a set of listed attack types: spoofing, tampering, repudiation, information disclosure, denial of service, and elevation of privilege (Hussain et al., 2014). However, there are also taxonomies and frameworks that deal with the threat actors. These include Intel’s TAL (Casey, 2007) and TARA (Rosenquist & Casey, 2009), SPEC (Gandhi et al., 2011), and Homeland Security Systems Engineering & Development Institute’s framework (Bodeau, McCollum & Fox, 2018). In 2015, Intel enhanced their model with motivational parameters (Casey, 2015). These frameworks do not list actual observed cyber adversaries but generate attack archetypes of attackers that differ according to their motivation.

We drafted a classification model for six threats based on motivational factors: cyber vandalism, cybercrime, cyber espionage, cyber terrorism, cyber sabotage, and cyber warfare. The motives can be reduced to their very essence: egoism, anarchy, money, destruction, paralysis, and power. The model was modified from Myriam Dunn Cavelty’s structural model (Dunn Cavelty, 2010; Ashenden, 2011; Lehto, 2013). Similar naming is used in multiple nations’ cyber security strategies (Luijff, Besseling & De Graaf, 2013), which increases usability of the model at different abstraction levels. In this publication the model creates the archetypes of cyber attacker used in scenario creation.

There is a need to describe the actual attackers and their actions behind archetypes. There are libraries of detected attacks and attacker groups. Mitre’s ATT&CK database (Strom et al., 2018) provides an updatable list of detected techniques, tools, and groups. The groups are tied to tactics, techniques, and tools they have been observed to use. However, Mitre’s database does not support the notion of motivation. In the SoS environment, examining only one actor is not sufficient (Bodeau & McCollum, 2018). Similarly, information about vulnerabilities forms sets of knowledge that can be used in conjunction with the protected system’s description to define the most likely paths of an attack. An example of vulnerability information database is the National Vulnerability Database (NVD; Booth, Rike & Witte, 2013).

However, all the knowledge about the actor, attack techniques, and vulnerabilities is irrelevant if it is not available where needed. This requires a precise shared terminology and a method for transferring the knowledge. The terminology should be able to address issues of varying abstraction levels and evolve with new concepts. Multiple information security companies have adopted Mitre’s ATT&CK as way to describe security events (Fireeye, 2021; Red Canary, 2021; ESET, 2019). ATT&CK also supports ICS and mobile system attack descriptions (Alexander, Belisle, & Steele, 2020; Strom et al., 2018).

Effective and secure sharing of information is also needed. STIX is considered the de facto threat information-sharing structure (Bodeau, McCollum & Fox, 2018), and TAXII is the method for creating sharing topologies. With STIX and TAXII, it is possible to create complex sharing networks that include the associated risk levels (Kokkonen

et al., 2016). This increases trust among participants and encourages sharing. Mitre’s ATT&CK is compatible with STIX (Bodeau, McCollum & Fox, 2018).

Understanding the motivation aspect enables prediction of situations that heighten the risk of a cyber-attack (Casey, 2015). When we combine the motivational factors for each attack archetype, we discover that different events trigger attacks. Many cyber-attacks are associated with social, political, economic, and cultural disputes, and the actions in the cyber domain may follow, precede, occur in parallel, or be uncorrelated with events in the physical world (Gandhi, 2011). Identifying the circumstances that might trigger an attacker archetype can be valuable in predicting heightened risk related to various situations.

The motivation affects the attacker’s targeting and methods. While a vandal seeks visibility by defacing a website, a spy wishes to stay unnoticed to gain information. The varying level of capability restricts some of the attackers from achieving their goals (Bodeau, McCollum & Fox, 2018). Therefore, having motivation does not mean that an attack is possible for the attacker. Understanding the motivations and capabilities of different archetypes limits the number of scenarios and thus makes evaluation feasible for the defender.

In the case of cyber vandalism, the arrival of a controversial vessel in a fairway might trigger actions. The controversy might be with the cargo, the vessel’s operations, or the vessel’s owner. For cybercrime, valuable cargo is more tempting as financial gains act as the motivation. Cyber espionage can include business or political espionage. Political factors may arise from national or international issues. From the national side, hacktivism supporting strikes in the harbor could be one scenario. In the worst case, international tensions in the region could escalate to military cyber operations against vessel traffic. The parameters of the attacker archetypes are presented in Table 1.

**Table 1:** Attributes of the attacker archetypes. Capability is derived from Bodeau, McCollum and Fox (2018), and impacts are derived from Mitre (2019a, 2020b)

	Vandalism	Crime	Espionage	Terrorism	Sabotage	Warfare operations
Motivation and goal	Trying to make political change based on personal political or ideological motives. Egoism gain	Making money through fraud or from the sale of valuable information. Financial gain	Gaining an economic, political, or military advantage. Information gain	Gaining social instability and influencing political decision-making. Anarchy gain	Causing instability, chaos, political change, and infrastructure paralysis. Paralysis gain	Performing a destructive attack on a nation’s digital infrastructure. Political or military dominance
Target	Digital services of governments and companies, individuals’ information systems	Digital services of governments and companies, individuals’ information systems	Data and information about governments and companies	Data and information about governments and companies. Critical infrastructure	Nation’s critical infrastructure	Nation’s critical infrastructure (civilian or military).
Capability	Acquired Attackers with moderate or limited expertise	Augmented Attackers with moderate or limited expertise	Advanced Attackers with very sophisticated or moderate expertise	Advanced Sophisticated attackers, capable of multiple, coordinated attacks	Integrated Very sophisticated attackers, capable of multiple, coordinated, continuous attacks	Integrated Very sophisticated attackers, capable of multiple, coordinated, continuous attacks
	Vandalism	Crime	Espionage	Terrorism	Sabotage	Warfare operations

	Vandalism	Crime	Espionage	Terrorism	Sabotage	Warfare operations
Trigger	A social event, an action of a company or an individual	The opportunity for economic gain	The need for political, economic, and military information	Cultural, nation's political or military actions	Testing own offensive cyber-attack capabilities, preparing hybrid or military operations	Achieving political or military objectives through military cyber operations
Impacts	ICT: Defacement ICT: Network Denial of Service ICS: Loss of Productivity and Revenue	ICT: Data Encrypted for Impact (ransom) ICT: Resource Hijacking (mining cryptocurrencies) ICS: Manipulation of Control (cargo capture at port)	ICS: Theft of operational information	ICS: Loss of Safety	ICS: Damage to Property (shipwreck)	ICS: Denial/Manipulation of View (GNSS attacks) ICS: Denial/Loss/Manipulation of Control (controlling ship) ICS: Damage to Property (shipwreck)

For ePilotage, the model of cyber-threat folds around the threat archetypes and their features. The actions they make are described with set vocabulary provided by ATT&CK. The support for non-trivial sharing schemas is essential in ePilotage SoS. This can be achieved by STIX and TAXII. With this information, drafting cyber-attack scenarios against ePilotage environment is possible.

### 3. Creating scenarios

Considering all the possibilities for attacks is resource intensive. Behind an attack, there are varying triggers depending on the attacker archetype. Examples of triggers that may motivate different actors and the impacts they are after are listed in Table 1. Evaluating all possible attack impacts will create too many scenarios for effective working. Evaluation and number of scenarios can be pruned with different archetypes and trigger events. For ePilotage, two different scenarios are created with different attacker archetypes. These scenarios demonstrate the methodology for constructing scenarios for ePilotage based on acquired knowledge.

Modeling threats in the SoS include challenges, such as risk governance, visibility, abstraction level, complexity, and external dependencies (Bodeau & McCollum, 2018). These relate to handling threat information with multiple stakeholders and systems. Attacker behavior does not change, but the targeting and effect propagation need more consideration. When Bodeau discusses SoS threat scenarios for the financial services sector, the SoS is larger than ePilotage's. For ePilotage, the main aspects to consider are possibilities for multiple attackers, common targetable technologies within SoS organizations, and key functionalities provided by one organization on which others depend. Moreover, supply chain attacks are a possibility.

#### 3.1 Cybercrime scenario

The first scenario involves cybercrime. A lot of information on cyber-crime trends, tools, and techniques is available publicly for example from cyber security companies' white papers. These are good starting points for the scenario. The motivation for this attack is gaining financial profit which can be achieved by selling individuals' personal information or stealing valuable cargo. Increasingly, cybercrime involves the use of ransomware and outsourcing malware distribution to Malware-as-a-Service (MaaS; Fireeye, 2020).

For example, Emotet is malware that was first used against financial services. It has been transformed into a MaaS platform delivering other payloads, including ransomware (Patsakis & Chrysanthou, 2020; Sophos, 2019). Mitre offers ATT&CK information on this malware and describes that it uses email phishing to spread (Mitre, 2019b). Ryuk is one ransomware that used Emotet with Trickbot for delivery (Cybereason, 2019).



For email to be an effective delivery mechanism, there must be a connection from the email recipient system to a system where the ransomware can cause damage that would make the victim willing to pay to restore the systems. Currently, pilot orders are accepted through email (Finnpilot, 2021). The recipient has very few reasons not to open a message that claims to be a pilot order. Emotet is known to have self-propagating features, such as spreading through server message block (SMB) shares. If it needs credentials, and it has brute force capabilities, which may accidentally lock out a valid account (Sophos, 2019). Moreover, Emotet uses the victim’s email contacts to send malicious email attachments to spread further (Sophos, 2019).

In ePilotage, this means the malware spreads in the SoS, because recent emails and fileshares most likely are used with other stakeholders. If Emotet is used to spread ransomware, a large portion of the SoS may become unusable. Even if the attack does not reach fairway systems or remote pilotage systems, investigation and recovery will take time. Meanwhile, the pilotage systems must be considered compromised. Mitre’s ATT&CK description and the suggested mitigations are in Table 2.

**Table 2:** Cyber-crime scenario with Mitre’s ATT&CK description of software, techniques, and mitigations (Mitre, 2021)

Software	ATT&CK technique	Mitigations
S0367 Emotet	T1566.001 Phishing: Spearphishing Attachment	Antivirus/Antimalware Network Intrusion Prevention Restrict Web-Based Content User Training
	T1204.002 User Execution: Malicious File	Execution Prevention User Training
	T1110.001 Brute Force: Password Guessing	Account Use Policies Multi-factor Authentication Password Policies
	T1078.003 Valid Accounts: Local Accounts	Password Policies Privileged Account Management
	T1021.002 Remote Services: SMB/Windows Admin Shares	Filter Network Traffic Limit Access to Resource Over Network Password Policies Privileged Account Management
	T1087.003 Account Discovery: Email Account	“This type of attack technique cannot be easily mitigated with preventive controls since it is based on the abuse of system features.” (Mitre, 2021)
S0446 Ryuk	T1486 Data Encrypted for Impact	Data Backup

### 3.2 Sabotage scenario

Creating an advanced cyber-attack scenario for an environment that does not exist is challenging. To make it more realistic, we need to state our assumptions and rely on reports of advanced attacks that have happened in other environments. Although zero days could be used, it should not be the magic word to pass through everything. For the scenario to be useful, there must be elements that can be detected and defended.

For a more advanced scenario, the SoS concerns can be studied in more detail. As the attacker’s capability increases, so does the campaign’s complexity and magnitude. As SoS attacks may target software, a service, or data that is used by multiple stakeholders, it is beneficial to choose such a target for this scenario. One of the targets might be common ePilotage software that holds a vulnerability, but at this phase of the design process, it is more feasible to choose another approach as the software is better evaluated later with more details. As a service, pilotage is at the center of ePilotage, and the data from the ships’ and fairway’s sensors is essential. If data is withheld, pilotage would cease.

There are multiple options for collecting data, such as sensors and data brokers or directly accessing the data storage. To develop the sabotage scenario, we make several assumptions. Sensors use the popular protocol of Message Queue Telemetry Transport (MQTT) and brokers to distribute data to subscribers. This protocol is susceptible to multiple security risks, including physical security and difficulties managing certificates (Perrone, 2017). In the SoS environment, it is likely that a needed portion of the data resides in the cloud for accessibility. A Denial of Service (DoS) attack is plausible in the cloud environment, but cloud providers often supply varying levels of protection against such attacks which would likely be easily detected. A successful DoS attack would

affect the availability of the data, but there are other possibilities. MQTT is a lightweight protocol but has risks if the sensors are not capable of handling encryption due to limited resources. If the sensors are widely dispersed in the fairway setting, their physical security is harder to organize leaving them vulnerable to physical access and attacks. If a sensor is affected, it can be used to attack the data broker. If the broker cannot be accessed due to the attack, the sensor data is lost. A validated example of MQTT broker DoS requiring low bandwidth is SlowITe (Vaccari, Aiello & Cambiaso, 2020). ePilotage operators would lose visibility to all sensors behind the affected broker. At first, they would not know whether the other sensors are affected. Likely, the operators would contact ships in the area to inform them about the technical issues and request sensor information from them. If the disturbance is not severe enough, the piloting would continue. At the minimum, the ships in the fairway would have to continue to a port or be directed elsewhere. A ship has its own sensors that send data to the ePilotage environment. This data partly covers the same topics as fairway sensors, such as the ship’s location.

This would open up the possibility for phishing attacks. Ships’ operators (remote or onboard) would have a reason to open an email concerning technical disruptions in the fairway that seem to come from a fairway staff member. Social engineering could cause some receivers to fulfill the order to “update” sent material to the Electronic Chart Display Information System (ECDIS) to secure it against disruptions. The ECDIS is not necessarily well protected or updated (PenTestPartners, 2018), and malware could change location information.

Another SoS attack issue is that an attack may have multiple targets. In this scenario, the adversary attacks a fairway sensor by conducting a DoS attack against the broker, denying access to the fairway sensor feed. Ships’ systems can be attacked through phishing attacks. This could be done to multiple ships in the area. The location of the fairway could be changed in the ECDIS causing at least one shipwreck. Simultaneously, VTS is flooded with contact requests based on contact information found online, resembling the customer service phone call flood in Ukraine in 2015 (Lee, Assante & Conway, 2016). As VTS communication records are available (VTS Finland, 2020), credible requests could be made or even automated.

The impacts of this type of attack are wide. Shipwreck(s) demand immediate actions to save lives. What will cause more trouble is restoring the operation of the fairway, as finding the cause and clearing the systems is not simple even in traditional ICT environments. A scenario utilizing Mitre’s ATT&CK and the suggested mitigations is in Table 3.

**Table 3.** Sabotage scenario with ATT&CK description of techniques and mitigations (Mitre, 2020a, 2021)

Target system	ATT&CK technique	Mitigations
MQTT data broker	ICS: T0856 Spoof Reporting Message (MQTT sensor hijacked physically)	Communication Authenticity Network Allowlists Software Process and Device Authentication Network Segmentation Filter Network Traffic
	ICS: T0815 Denial of View (MQTT broker unable to handle sensor’s spoofed messages)	Out-of-Band Communications Channel Redundancy of Service Data Backup
Ship	ICT: T1589.002 Gather Victim Identity Information: Email Addresses	“This technique cannot be easily mitigated with preventive controls since it is based on behaviors performed outside of the scope of enterprise defenses and controls. Efforts should focus on minimizing the amount and sensitivity of data available to external parties.”
	ICT: T1566.001 Phishing: Spearphishing Attachment (request to immediately apply critical update sent to the ECDIS due to current technical issues)	Antivirus/Antimalware Network Intrusion Prevention Restrict Web-Based Content User Training
	ICS: T0847 Replication Through Removable Media (ECDIS USB for traditional ships)	Disable or Remove Feature or Program Limit Hardware Installation Operating System Configuration

Target system	ATT&CK technique	Mitigations
	ICS: T0822 External Remote Services (Remote update for automated ships)	Disable or Remove Feature or Program Multi-factor Authentication Network Segmentation User Account Management Limit Access to Resource Over Network Account Use Policies Password Policies
	ICS: T0863 User Execution (updating ECDIS)	Antivirus/Antimalware Code Signing Execution Prevention Network Intrusion Prevention Restrict Web-Based Content User Training
	ICS: T0875 Change Program State (malware takes control)	Authorization Enforcement Human User Authentication Communication Authenticity Network Allowlists Access Management Software Process and Device Authentication Network Segmentation Filter Network Traffic
	ICS: T0836 Modify Parameter (fairway parameters changed)	Authorization Enforcement Audit
	ICS: T0832 Manipulation of View (false parameters shown)	Communication Authenticity Out-of-Band Communications Channel Data Backup
	ICS: T0879 Damage to Property (shipwreck)	Network Allowlists Mechanical Protection Layers Safety Instrumented Systems
VTS	ICT: T1591 Gather Victim Org Information (VTS contact information and contact details)	“This technique cannot be easily mitigated with preventive controls since it is based on behaviors performed outside of the scope of enterprise defenses and controls. Efforts should focus on minimizing the amount and sensitivity of data available to external parties.”
	ICT: T1498 Network Denial of Service (by connection requests to VTS operators)	Filter Network Traffic

#### 4. Scenario analysis

A scenario can tell one story of what might happen. It provides material for creating test cases and tabletop exercises, but it does not present everything that could happen. In the sabotage case, there would have been opportunities for more direct attacks, such as a DoS to the cloud or assuming a direct internet connection to the ICS. In an examination of Mitre’s ICS cases, valid accounts gained through phishing or drive-by-download were the most common route to controlling ICS (Mitre, 2020a). In these cases, there is no need for advanced malware or complicated routes for exploitation. Using legitimate commands with a valid account is sometimes enough. The more complicated scenarios aid in revealing dependencies between systems and events.

Creating the scenarios showed that a versatile knowledge database can help utilize current malware campaigns in scenarios. Previous attacks and demonstrations of proof-of-concepts create validation for the steps of the attacks. Another benefit of adapting real cases to ePilotage is the possibility for scenario users to find additional deeper technical details of the attacks. Scenarios using specific malware or vulnerabilities quickly become outdated. For example, a joint effort by law enforcement and other authorities disrupted the Emotet infrastructure in an operation in January 2021 (Europol, 2021). However, this does not render the scenario obsolete, as ransomware can be delivered through another MaaS. This was seen when ransomware Ryuk, which used Emotet for distribution, was observed shifting to a different MaaS provider, called Buer Loader (Callagher, 2020).

Mitre’s database suggestions for mitigations cannot be followed blindly because not all mitigations are possible in ePilotage. For example, removing the USB update option from all current vessels’ ECDIS is not feasible. Even

these short scenarios produced 31 different mitigation options, and thus, security does not have only one solution. Many of the mitigations relate to knowing who is operating a system and restricting usage with allowlists and segmentation. Good passwords and trained users also get to the top of the list. Achieving this in a multiorganization environment, and controlling that everyone follows it, is not easy. The same challenge arises from hardening systems of which some are not maintained within ePilotage.

The popularity of ATT&CK's helps integrate cyber-security reports and white papers more easily in created scenarios as material in the correct format exists. Using Mitre's ATT&CK proved possible in ePilotage even when the scenario included ICS environments. The VTS DoS was the most vaguely described by ATT&CK terminology as the attack does not use the traditional ICT network. The tables provide concise descriptions of the scenarios with explicit terminology. Furthermore, readers are given the possibility to increase their knowledge by examining other attacks that utilized the same techniques. However, ATT&CK does not support the attacker archetypes or motivational factors. These must be included in the scenario description and are more often used in the first phases of scenario creation. At the beginning, when the targets and desired impacts are planned, these aspects are more important. While using the ATT&CK, the archetypes aid in selecting the proper level of capability used by the actors.

## **5. Conclusions and future work**

The analysis of the two scenarios revealed that ePilotage and maritime traffic face similar vulnerabilities. Attack propagation can differ when transferred to remotely controlled and automated traffic, and security mechanisms will likely improve over time. The dependency on sensor information will increase when the human element onboard is removed. Deciding which information to trust in a situation when something unexpected happens is not easier when inspections of shoreline locations cannot be confirmed by simply looking. As time is of the essence in a space-constrained environment, such as a narrow fairway path, the situational awareness needs to be good and the communication precise.

Future work will consist of refining the scenarios with a more precise technical description of the ePilotage environment and then playing tabletop exercises to develop a more robust environment. Moreover, defining the shared topologies of threat information through STIX and TAXII requires designing for ePilotage needs.

## **References**

- Alexander, O., Belisle, M., and Steele, J. (2020) "Mitre ATT&CK® for Industrial Control Systems: Design and Philosophy".
- Ashenden, D. (2011) "Cyber Security: Time for Engagement and Debate", Proceedings of the 10th European Conference on Information Warfare and Security, the Institute of Cybernetics at the Tallinn University of Technology.
- Bodeau, D.J. and McCollum, C.D. (2018) "System-of-systems Threat Model", The Homeland Security Systems Engineering and Development Institute (HSEDI) MITRE, Bedford.
- Bodeau, D.J., McCollum, C.D. and Fox, D.B. (2018) "Cyber Threat Modeling: Survey, Assessment, and Representative Framework", Mitre Corp, Mclean.
- Booth, H., Rike, D. and Witte, G.A. (2013) "The National Vulnerability Database (NVD): Overview", ITL BULLETIN, Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, U.S. Department of Commerce.
- Callagher, S. (2020) "Hacks for sale: inside the Buer Loader malware-as-a-service", [online], <https://news.sophos.com/en-us/2020/10/28/hacks-for-sale-inside-the-buer-loader-malware-as-a-service/>.
- Casey, T. (2007) "Threat Agent Library Helps Identify Information Security Risks", Intel White Paper, p 2.
- Casey, T. (2015) "Understanding Cyber Threat Motivations to Improve Defense", Intel White Paper.
- Cybereason (2019) A One-two Punch of Emotet, TrickBot, & Ryuk Stealing & Ransoming Data, [online], <https://www.cybereason.com/blog/one-two-punch-emotet-trickbot-and-ryuk-steal-then-ransom-data>.
- DHS. (2013) "National Infrastructure Protection Plan (NIPP) - Partnering for Critical Infrastructure Security and Resilience".
- DIMECC Oy (2020) "Sea4Value/Fairway Program (S4VF)", [online], <https://www.dimecc.com/dimecc-services/s4v/>.
- Dunn Caveltly, M. (2010) "The Reality and Future of Cyberwar", Parliamentary Brief, 30 March 2010, [online], [www.parliamentarybrief.com/2010/03/the-reality-and-future-of-cyberwar](http://www.parliamentarybrief.com/2010/03/the-reality-and-future-of-cyberwar).
- ESET. (2019) "The MITRE ATT&CK Framework: Everything You Need to Know in Under 60 Minutes", [online], <https://www.eset.com/us/business/resources/webinars/the-mitre-attck-framework-everything-you-need-to-know-in-under-60-minutes-1/>.
- Europol. (2021) "World's Most Dangerous Malware EMOTET Disrupted Through Global Action", [online], <https://www.europol.europa.eu/newsroom/news/world%E2%80%99s-most-dangerous-malware-emotet-disrupted-through-global-action>.
- Finnpilot (2021) "Tilaa luotsaus/Pilot Online", [online], <https://finnpilot.fi/asiakkaille/tilaa-luotsaus-ja-pilot-online/>.
- Fireeye (2020) "M-Trends 2020". Fireeye Mandiant Services, Special Report.

***Tiina Kovanen, Jouni Pöyhönen and Martti Lehto***

- Gandhi, R., Sharma, A., Mahoney, W., Sousan, W., Zhu, Q. and Laplante, P. (2011) "Dimensions of Cyber-Attacks: Cultural, Social, Economic, and Political", IEEE Technology and Society Magazine, Vol. 30, No. 1, pp 28–38.
- Hussain, S., Kamal, A., Ahmad, S., Rasool, G. and Iqbal, S. (2014) "Threat Modelling Methodologies: A Survey", Science International (Lahore), Vol. 26, No. 4, pp 1607–1609.
- Kokkonen, T., Hautamäki, J., Siltanen, J. and Hämäläinen, T. (2016) "Model for Sharing the Information of Cyber Security Situation Awareness Between Organizations", in 2016 23rd International Conference on Telecommunications (ICT), IEEE, pp 1–5.
- Kovanen, T., Pöyhönen, J. and Lehto, M. (2021) "ePilotage System of Systems' Cyber Threat Impact Evaluation", accepted for ICCWS21.
- Lee, R., Assante, M. and Conway, T. (2016) "Analysis of the Cyber Attack on the Ukrainian Power Grid", Defence Use Case. E-ISAC.
- Lehto, M. (2013) "The Cyberspace Threats and Cyber Security Objectives in the Cyber Security Strategies", International Journal of Cyber Warfare and Terrorism, Vol. 3, No. 3, pp 1–18.
- Luijff, E., Besseling, K. and De Graaf, P. (2013) "Nineteen National Cyber Security Strategies", International Journal of Critical Infrastructures 6, Vol. 9, No. 1–2, pp 3–31.
- Mavroeidis, V. and Bromander, S. (2017) "Cyber Threat Intelligence Model: An Evaluation Of Taxonomies, Sharing Standards, and Ontologies Within Cyber Threat Intelligence", in 2017 European Intelligence and Security Informatics Conference (EISIC), IEEE, pp 91–98.
- Mitre. (2019a) "Impact", [online], <https://attack.mitre.org/tactics/TA0040/>.
- Mitre. (2019b) "Emotet", [online], <https://attack.mitre.org/software/S0367/>.
- Mitre. (2020a) "ATT&CK® for Industrial Control Systems", [online], [https://collaborate.mitre.org/attackics/index.php/ImpactMain\\_Page](https://collaborate.mitre.org/attackics/index.php/ImpactMain_Page).
- Mitre. (2020b) "Impact", [online], <https://collaborate.mitre.org/attackics/index.php/Impact>.
- Mitre. (2021) "ATT&CK Matrix for Enterprise", [online], <https://attack.mitre.org/>.
- Patsakis, C. and Chrysanthou, A. (2020) "Analysing the Fall 2020 Emotet Campaign", [online], arXiv preprint arXiv:2011.06479.
- PenTestPartners. (2018) "Hacking Serial Networks on Ships", [online], <https://www.pentestpartners.com/security-blog/hacking-serial-networks-on-ships/>.
- Perrone, G., Vecchio, M., Pecori, R. and Giaffreda, R. (2017, April) "The Day After Mirai: A Survey on MQTT Security Solutions After the Largest Cyber-attack Carried Out Through an Army of IoT Devices", in IoTBDS, pp 246–253.
- Pöyhönen, J., Kovanen, T. and Lehto, M. (2021) "The Basic Elements of Cyber Security for Automated Remote Piloting Fairway System", accepted for ICCWS 2021.
- Red Canary. (2021) "2020 Threat Detection Report", [online], <https://redcanary.com/threat-detection-report/>.
- Rosenquist, M. and Casey, T. (2009) "Prioritizing Information Security Risks With Threat Agent Risk Assessment", Intel Corporation White Paper.
- Sophos (2019) "Emotet Exposed: Looking Inside Highly Destructive Malware", Network Security, Vol. 2019, No. 6, pp 6–11.
- Strom, B.E., Applebaum, A., Miller, D.P., Nickels, K.C., Pennington, A.G. and Thomas, C.B. (2018) "Mitre Att&ck: Design and Philosophy", Technical report.
- Vaccari, I., Aiello, M. and Cambiaso, E. (2020) "SlowITe, a Novel Denial of Service Attack Affecting MQTT", Sensors, Vol. 20, No. 10, p 2932.
- VTS Finland (2020) "Master's Guide Vessel Traffic Services. Helsinki VTS, Sector 1 (7.2.2020)", [online], <https://tmfg.fi/sites/default/files/2020-02/Helsinki%20VTS%20Sector%201%20EN.pdf>.
- Xiong, W. and Lagerström, R. (2019) "Threat Modeling—A Systematic Literature Review", Computers & Security, Vol. 84, pp 53–69.

# Impact of AI Regulations on Cybersecurity Practitioners

Louise Leenen<sup>1,2</sup>, Trishana Ramluckan<sup>3,4</sup> and Brett van Niekerk<sup>3</sup>

<sup>1</sup>University of the Western Cape, South Africa

<sup>2</sup>Centre for AI Research, South Africa

<sup>3</sup>University of Kwazulu-Natal, South Africa

<sup>4</sup>Educor Holdings, South Africa

[lleenen@uwc.ac.za](mailto:lleenen@uwc.ac.za)

[trishana.ramluckan@educor.co.za](mailto:trishana.ramluckan@educor.co.za)

[vanniekerkb@ukzn.ac.za](mailto:vanniekerkb@ukzn.ac.za)

DOI: 10.34190/EWS.21.014

**Abstract:** Cybersecurity and Artificial Intelligence (AI) are closely aligned; both domains are growing rapidly, they are inter-dependent and they face major challenges. Cybersecurity threats are of great concern globally and becoming increasingly difficult to address. AI is widely recognised to be crucial to the development of efficient cybersecurity measures. AI is a core tool in combatting cybercrime and other cybersecurity threats but many AI applications are, in turn, vulnerable to cybersecurity attacks. Most countries have some cybersecurity and privacy legislation in place or are in the process of putting regulations in place. However, AI regulation is still relatively new. Due to the close relationship between cybersecurity and AI, cybersecurity practitioners have to be aware of national, regional and international AI regulations. In this paper, an overview of the progress in terms of AI governance and regulations is provided, including national AI policies and international agreements on AI issues. The paper concludes with recommendations for the cybersecurity community.

**Keywords:** artificial intelligence regulations, cybersecurity governance, AI principles, ethics in AI

---

## 1. Introduction

The rapid global growth in cybersecurity threats is matched by an equally rapid growth in Artificial Intelligence (AI) applications to build counter-measures against these threats. AI is widely recognised to be crucial to the development of efficient cybersecurity measures (Columbus, 2020) (Dua & Du, 2011). The European Union Agency for Cybersecurity (ENISA) published a report in December 2020, *AI Security Challenges: Threat Landscape for AI*, in which it is noted that “Artificial Intelligence and cybersecurity have a multi-dimensional relationship and a series of interdependencies”. Cybersecurity can be one of the foundations of trustworthy Artificial Intelligence solutions (ENISA, 2020). In its report, ENISA issues a warning that “AI may open new avenues in manipulation and attack methods, as well as new privacy and data protection challenges”. They advise that there should be a common understanding of the relevant threat landscape and associated challenges. The report highlights the need for a European Union (EU) ecosystem for secure and trustworthy AI that includes all elements relating to the AI supply chain.

AI clearly has an impact on cybersecurity, but cybersecurity also influences AI. AI techniques and implementations are increasingly used to counter cyber-attacks, but AI applications can also be vulnerable to cyber-attacks and be employed by cyber criminals to launch cyber-attacks (Leenen & Meyer, 2019). One example is the hacking of self-driving vehicles (Lemos, 2020). Dreyfuss (2019) discusses reports by a hacker that claims to have hacked into 27 000 GPS tracking accounts and gained control to the vehicles’ internal systems. The hacker said he was able to track all the cars and turn off the engines of cars under certain conditions. This breach was independently verified. IBM stated that “As cyberattacks grow in volume and complexity, artificial intelligence (AI) is helping under-resourced security operations analysts stay ahead of threats” (IMB, n.d.) AI is becoming more prominent in many applications areas, for example, in advancing automation. However, AI also poses risks to society; loss of privacy, exploitation of underlying biases in data, loss of jobs, etc. It is evident that AI has to be regulated to produce applications that are accountable, responsible, explainable and ethical.

The societal impact of these two domains has led to a recognition of a need for governance and regulations. Whilst a lot of attention has been paid to cybersecurity regulations and legislation globally (United Nations Conference on Trade and Development (UNCTAD), n.d.), AI regulation is still in an early phase of development. The interconnectedness of these two domains requires a holistic approach to governance and regulation measures. Because societies can be adversely affected by cybersecurity threats and by AI applications, it is vital for researchers, practitioners, and policy makers to guide the development and coordinate interventions so that all stakeholders are considered during the development of these measures. The cybersecurity community will

have to fully embrace AI in order to keep up with the rapid growth in cybersecurity threats. It is important that cybersecurity practitioners have to be aware and understand the implications of not just cybersecurity regulations but also AI regulations. This paper presents a view on the impact AI regulations is likely to have on the cybersecurity domain and provides some recommendations. The authors believe that the development of AI regulations and governance is likely to follow the same path as that of cybersecurity regulations and governance. The next subsection provides a brief overview of the status of cybersecurity regulations. Section 2 explores the strong connection between the two disciplines, cybersecurity and AI. Several countries have formulated, or are in the process of, formulating national AI Policies and regulations. Section 3 gives an overview of AI regulations. Section 4 considers the impact of AI regulations on cybersecurity practitioners and provide some recommendations for cybersecurity practitioners. The paper is concluded in Section 5.

### 1.1 Cybersecurity regulations and governance

Cybersecurity is often regulated and governed at a national level through acts and national strategies. These typically relate to privacy, defining cyber-crime and establishing punishments for these crimes, and for protecting critical national infrastructure. There are some regional initiatives, such as the *South African Development Community Model Law* (International Telecommunications Union, 2013) and the *European Union's Cybersecurity Act* (European Union, 2020). These regional documents aim at providing consistent regulation across a region to provide a degree of confidence that the nations in a region will have compatible laws to facilitate governance across borders. Where cybersecurity regulations and governance begin to struggle is at the international level. *The Budapest Convention* (Council of Europe, 2001) is an attempt to have a common understanding and allow for collaboration against cyber-crime. However, challenges still exist. While the United Nation's Group of Governmental Experts (2015) confirmed that international humanitarian law applies to cyberspace, the challenge is in translating the existing laws to the digital realm. The efforts usually take the form of academic studies, such as the *Tallinn Manual 2.0* (Schmitt, 2017), or voluntary norms, such as the *Global Commission for the Stability of Cyberspace's Report* (2019) and *The Paris Call on Trust and Security in Cyberspace* (2018). Some nations, such as the Netherlands and France, have made statements on their interpretation of the applicability of international law to cyberspace (van Niekerk, Ramluckan, & Ventre, 2020). Figure 1 shows the commitment of countries to cyber (International Telecommunications Union (ITU), 2018).

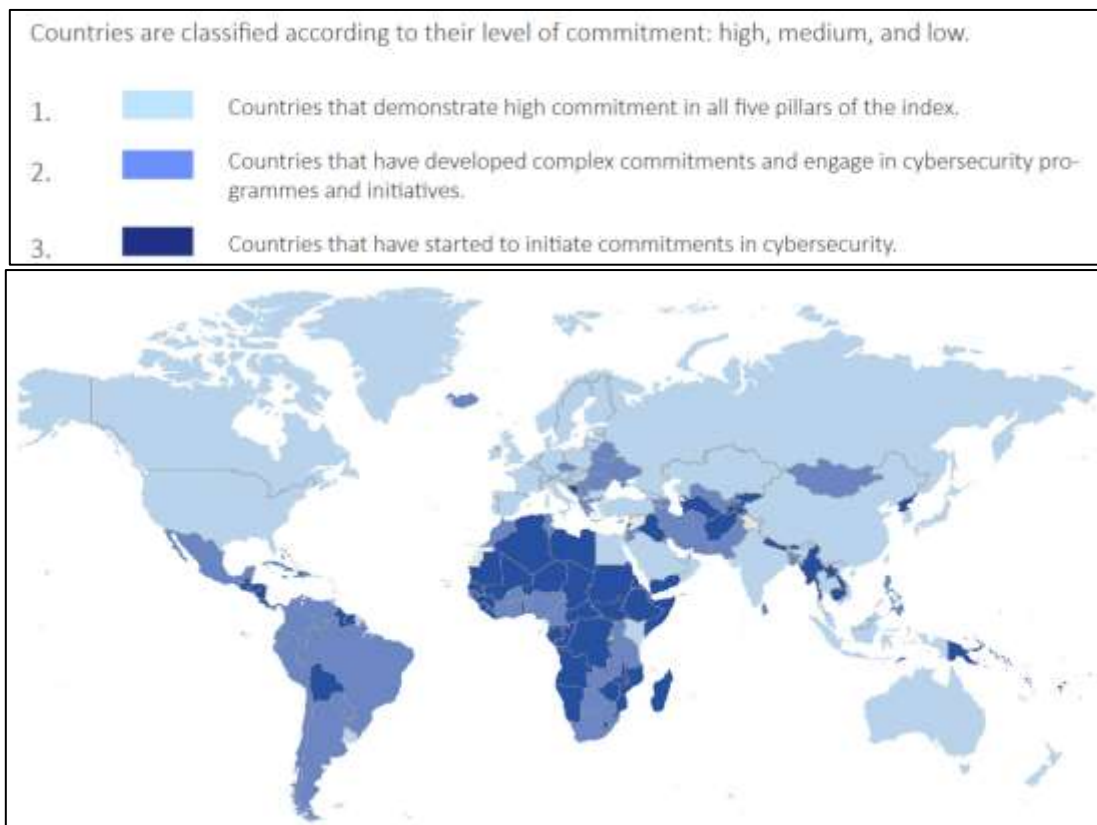


Figure 1: Global cyber commitment (International Telecommunications Union (ITU), 2018)

While the laws are set at national levels, with some international collaboration or alignment, there needs to be technical implementation of cybersecurity that is governed at an organisational or infrastructure provider level. These often take the form of frameworks or standards, such as the *ISO/IEC 27000-series standards*, the *Top 20 Cyber Security Controls* (Centre for Internet Security, 2019), and the *NIST Cyber Security Framework* (NIST, 2013). Such documents cover the implementation, management and governance of preventative measure, detection, and response and recovery. By implementing such frameworks and/or standards, organisations can then achieve objectives and requirements laid out by national and/or regional regulations.

## **2. Interdependency of cybersecurity and artificial intelligence**

There is a growing interdependency between the AI and cybersecurity domains. In this section, we discuss the use of AI techniques in cyber counter-attack systems, cybersecurity for AI systems and the malicious use of AI in developing and launching cyber-attacks.

ENISA's report on Cybersecurity and AI focuses on three key dimensions: AI to support or facilitate cybersecurity, the malicious use of AI to conduct cyber-attacks, and cybersecurity to protect the possible vulnerabilities in AI (ENISA, 2020). In order to aid cybersecurity, AI can be used to automate a number of processes which will be difficult for humans, such for network threat detection, email scanning, and attack or malware classification (Ye, Cheng, Zhu, Feng, & Song, 2018), (Yerima, Sezer, & Muttik, 2015), (ENISA, 2020). Some solutions using AI have been commercially implemented, such as the products for DarkTrace (2021) and Deep Instinct (2020). There are also possibilities for enhancing existing security controls such as 'smart' forensics and active or adaptive firewalls (ENISA, 2020).

However, the malicious use of AI is also growing. Generative Adversarial Networks have been used to create deep-fakes that were successfully used in advanced scams (Damiani, 2019), and AI has been used to control botnets to successfully attack an online marketplace (Bocetta, 2020). It is predicted that the use of AI in cyber-attacks will increase, with the potential for adaptive cyber-attacks that can learn to avoid detection (Heinemeyer, 2020). There is also the possibility of AI controlled social media bots, controlling compromised social media accounts to broadcast misinformation and propaganda, similar to the use of bots during the COVID pandemic (Allen-Ebrahimian, 2020), (Kao & Li, 2020). This could then be extended to AI-augmented DDoS attacks and AI-supported password cracking (ENISA, 2020). In addition to the deployment of AI for malicious purposes, AI implementations could be subverted (ENISA, 2020). Attempts to target an AI system could include data poisoning at various stages that could affect training, model validation, feature identification, denial-of-service, or result in other spurious outputs (ENISA, 2020). An example is modified stop signs resulted in the AI of automated cars interpreting the sign as a speed sign (Field, 2017).

Given the possibility of subverting AI implementations, there is a need to provide cybersecurity solutions to protect AI implementations (ENISA, 2020). In particular, many AI systems require data which needs to be stored and transported, as well as processed by the AI algorithms. This necessitates data security and privacy controls to protect that data. As indicated above, data poisoning may impact various stages of the AI supply chain, therefore the integrity of data needs to be ensured at all stages to have confidence in the model and its performance. Such concerns may also introduce the need for more rigorous testing of AI systems through the use of fuzzing, where random or modified inputs can be provided to ascertain if irregular outputs occur.

Another sub-domain in which AI is proving to be crucial is to support compliance with privacy and data protection legislation. Article 5 of the *General Data Protection Regulation* (GDPR) establishes security as a principle when processing personal data (Information Commissioner's Office, n.d.). This principle is a significant shift that elevates the role of security from its previous role as a technical provision. The GDPR makes security a pre-requisite, and a lack of appropriate security measures makes the processing of sensitive data unlawful. Gartner Inc predicts by 2023 more than 40% of privacy compliance technology will rely on AI, compared to only 5% in 2020 (Gartner, 2020). More than 60 countries and regions have introduced, or are in the process of introducing, privacy and data protection legislation. Gartner explains this considerable increasing reliance on AI technology is in a large part due to subject rights requests (SRRs). An SRR covers a set of rights an individual has to information on his or her data, and organisations are obliged to respond to such requests within a defined period of time. According to the *2019 Gartner Security and Risk Survey*, two third of organisations are unable to respond swiftly to SRRs, and others are doing this manually which is very expensive. AI-based tools can handle large

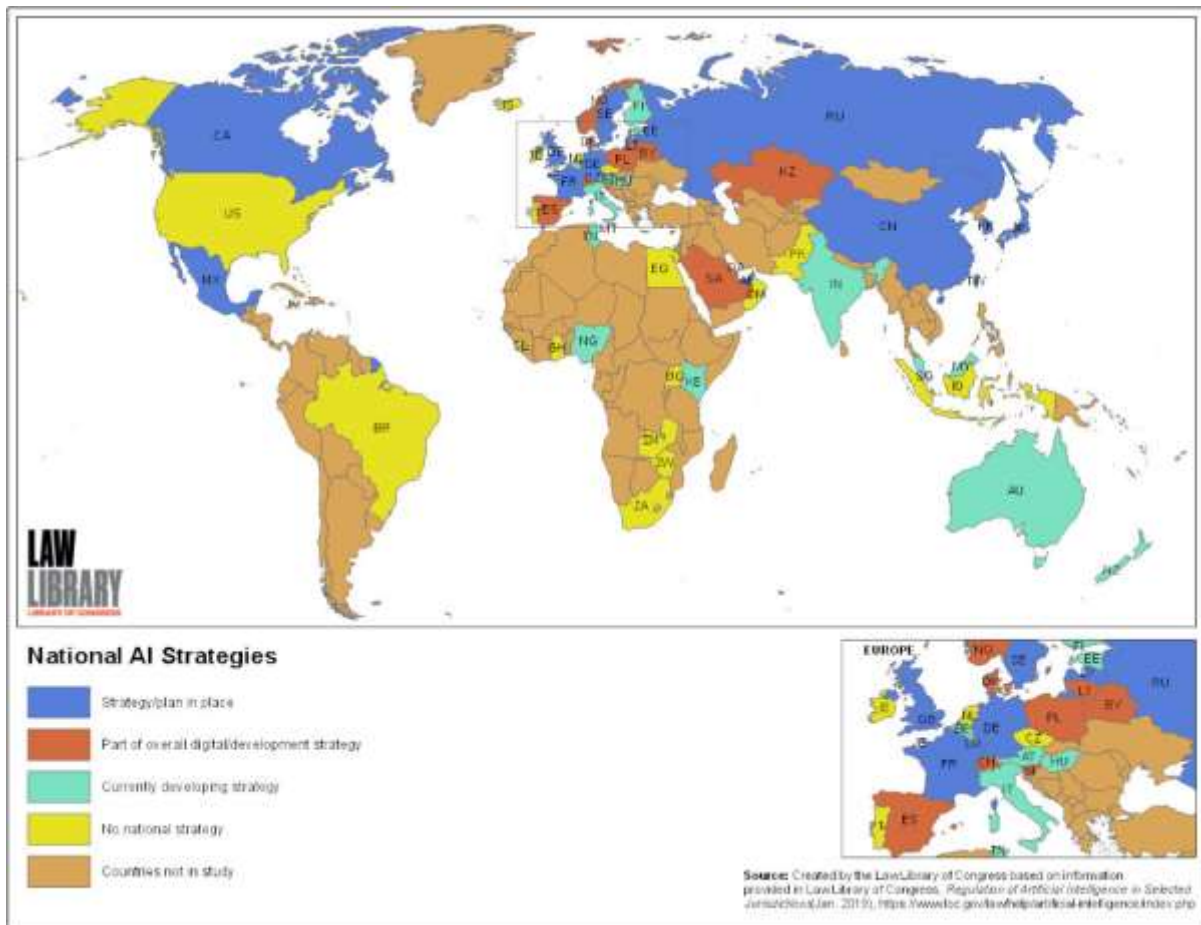


volumes of SRRs fast and thus reduce costs and build customer trust. AI also support other privacy compliance processes. (Gartner, 2020).

### 3. Overview of artificial intelligence policies and regulations

The regulation of AI is an emerging domain. There is recognition that AI applications can affect society negatively and that regulations based on commonly agreed upon principles are required. Some countries and international bodies have made some progress in this regard, but many countries do not have any regulations in place as yet.

The *Regulation of Artificial Intelligence in Selected Jurisdictions* (The Law Library of Congress, Global Legal Research Directorate., 2019) report, published by the United States Government in 2019, reviews AI regulations, strategies and policies in several countries and in the European Union (EU). This survey is valuable in an emerging domain; it describes different approaches to AI governance, including those followed by of international organisations such as the United Nations (UN) and other institutions. Figure 2 is copied from this report and shows the progress globally in establishing AI regulations.



**Figure 2:** National AI strategies taken from the *Regulation of Artificial Intelligence in Selected Jurisdictions* Report (The Law Library of Congress, Global Legal Research Directorate., 2019)

The Oxford Insights institution and the International Research Development Centre (based in Canada) have jointly published AI Readiness indices for governments over a period of three years, resulting in the *Government AI Readiness 2020* report (Oxford Insights and the International Research Development Centre, 2020). Section 3.2. discusses this report.

The report, *Examples of AI National Policies* (Organisation for Economic Co-operation and Development, 2020) published by the OECD, served as input for a discussion amongst members of the G20 Digital Economy Task Force in 2020. The Council of Europe, the European Parliament (European Parliament, 2019), the G7 countries, the OECD (Organisation for Economic Co-operation and Development, 2019) have had discussions on AI governance and how to deal with Big Data (Twomey, 2020). The G20 Trade Ministers and Digital Economy Ministers adopted a set of AI Principles (G20, 2019) in 2019, referred to as the *G20 AI Principles*. These Principles

are to a large extent based on the OECD's principals and discussions held by G20 groups. A set of commonly agreed upon principles among a large set of countries and institutions is an invaluable foundation for global cooperation even if they are relatively high-level statements. It provides a starting point for more detailed regional and international agreements. An overview of the *G20 AI Principles* is given in Section 3.1, and in Section 3.2 an overview of national AI policies and developments in a few countries are presented.

### **3.1 G20 AI principles**

The G20's principles (G20, 2019) prescribe responsible stewardship of trustworthy AI (Part 1) and recommendations for national policies and international co-operation for trustworthy AI (Part 2). The principles prescribe an ethical and human-centred approach to guiding AI governance and strategy formulation, and provide a good basis for establishing trust and cooperation in the future development of AI.

Part 1 contains five principles:

- Inclusive growth, sustainable development and well-being;
- Human-centered values and fairness;
- Transparency and explainability;
- Robustness, security and safety; and
- Accountability.

Part 2 contains the following recommendations for national policies:

- Investing in AI research and development;
- Fostering a digital ecosystem for AI;
- Shaping an enabling policy environment for AI;
- Building human capacity and preparing for labour market transformation; and
- International co-operation for trustworthy.

### **3.2 Progress in developing AI regulations**

The *Government AI Readiness 2020* report (Oxford Insights and the International Research Development Centre, 2020) considers the readiness level of a given government to implement AI in order to provide public services to their citizens. AI can be used to improve the delivery of healthcare, education, transport and many other services. The authors of this report divided the world in 9 regions and selected 2 to 3 countries from each region for a more detailed overview. The United States of America (USA) is the top ranked country. Western Europe scores high in terms of the number of countries with National AI Policies in place and having a regional strategy outlined in a 2020 white paper, *European approach to excellence and trust* (European Commission, 2020). The UK, Finland, Germany and Sweden scores 2<sup>nd</sup> to 5<sup>th</sup> highest after the USA. China is ranked at position 19, the Russian Federation at position 33, India at position 40, South Africa at position 59, and Brazil at position 63. Mauritius is the highest ranked Sub-Saharan African at position 45. It is worth noting that this report also provides a sub-index for Responsible Use of AI. The *Global Partnership on Artificial Intelligence* (Global Partnership on Artificial Intelligence, 2020) should also be noted; Australia, Canada, France, Germany, India, Italy, Japan, Mexico, New Zealand, the Republic of Korea, Singapore, Slovenia, the United Kingdom, the United States of America, and the European Union joined this initiative.

The G20's *Examples of AI National Policies* report (Organisation for Economic Co-operation and Development, 2020) provide overviews of the progress of AI regulations and legislation in several countries. Some summaries are extracted below:

*The USA:* The National Institute of Standards and Technology (NIST) published *U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools* in August 2019 (NIST, 2019) which advises U.S. government agencies on AI regulations efforts. In 2020, the White House's Office of Management and Budget (OMB) published the draft memorandum, *Guidance for Regulation of AI Applications* (OMB, 2020), which is strongly aligned to the values-based G20 AI Principles.

*Germany:* Germany adopted an AI strategy in 2018 that includes goals and steps to ensure Germany is a leading force in AI development (Bundesregierung, 2018). It has a focus in incorporating society in AI issues and addresses ethics and responsible use of AI.

*Brazil:* Brazil's National AI Strategy, *Estratégia Brasileira de Inteligência Artificial*, (Participa.br, 2020) is currently open for public comment. The strategy is based on a whole-of-society approach to AI and aims to use AI for scientific development, competitiveness and productivity (including in public services) and well-being. It adopts all the G20 AI Principles.

*China:* China's *Governance Principles for the New Generation AI* released in 2017 (Roberts, et al., 2020) promotes the responsible development of AI and span the five values-based G20 AI Principles. China also has a draft *Artificial Intelligence Industry Alliance* (AIIA) that contains most of G20 principles.

*India's AI Strategy:* India is focusing on leveraging AI for inclusive growth. The Indian AI Strategy and its implementation align strongly with many aspects of the G20 AI Principles. In 2018, India established a *National Programme on AI9* (Niti Aayog, 2018) to guide Research and Development in new and emerging technologies.

*The Russian Federation:* In October 2019, Russia adopted the *National Strategy for AI Development* (Kremlin, 2019) to serve as the basis for development and enhancement of state programmes and projects. The strategy has a strong focus on ethics and addresses a number of the values-based G20 AI Principles. In addition, measures for fostering of a digital ecosystem for AI are supported.

*South Africa:* South Africa does not yet have a national AI policy in place. A *Presidential Commission on the Fourth Industrial Revolution* (4IR) was formed in 2019 (South African Government, 2020). In October 2020, the Minister Communications and Digital Technologies Minister announced the release of the Commission's report which is the first step towards a national policy.

#### **4. The Impact of AI regulations on cybersecurity practitioners**

Globally, the development of National AI Policies and Strategies is in an early level of maturation, and in many regions it is still in its infancy. Cybersecurity governance and regulations is at a more mature level and the cybersecurity community has invested in raising awareness amongst its practitioners. However, it is indisputable that AI and cybersecurity are interdependent and there is a need for the cybersecurity community to become aware of the emerging consequences. The AI domain is growing fast and changes constantly, and this will result in subsequent and ongoing changes to AI principles of international and regional bodies and AI regulations. Regulations should thus be adjusted to make provisions for the impact that emerging technologies and applications will have; examples are autonomous weapons and self-driving cars – although there are countries that already have regulations place for these applications, there is concern about the uncertainty regarding the future impact they may have. For example, Figure 3 illustrates the current approaches to regulation of autonomous weapons. The consequence is that raising awareness amongst cybersecurity practitioners of AI regulations will also be an ongoing activity.

A difficult issue to resolve is the role that country specific AI regulations will have in a globally connect world. Pomares and Abdala (Pomares & Abdala, 2020) discuss this issue at length and conclude that although there are many differences in national approaches, commonly accepted AI Principles provide a good starting point to a complex process that needs to take place to establish globally accepted regulations. The development of regional cybersecurity regulation and legislation (section 4.1.) have matured somewhat compared to AI regulation and legislation. It takes time to develop regional and international collaborations to produce commonly agreed upon measures.

It has become apparent that privacy and data protection legislation, and AI regulations and subsequent legislation, will regard security as an integral aspect of AI applications - developers will be held responsible and accountable for the security of their products. This is likely to have a significant impact on cybersecurity practitioners and they should be trained and consistently informed about any relevant regulations or legislation. Cybersecurity practitioners need to be familiar with a number of facets of computing, such as programming, networking, operating systems and architecture; with the growing prevalence of AI implementations and security concerns related to AI, this is another area that security professionals will need to be familiar with.

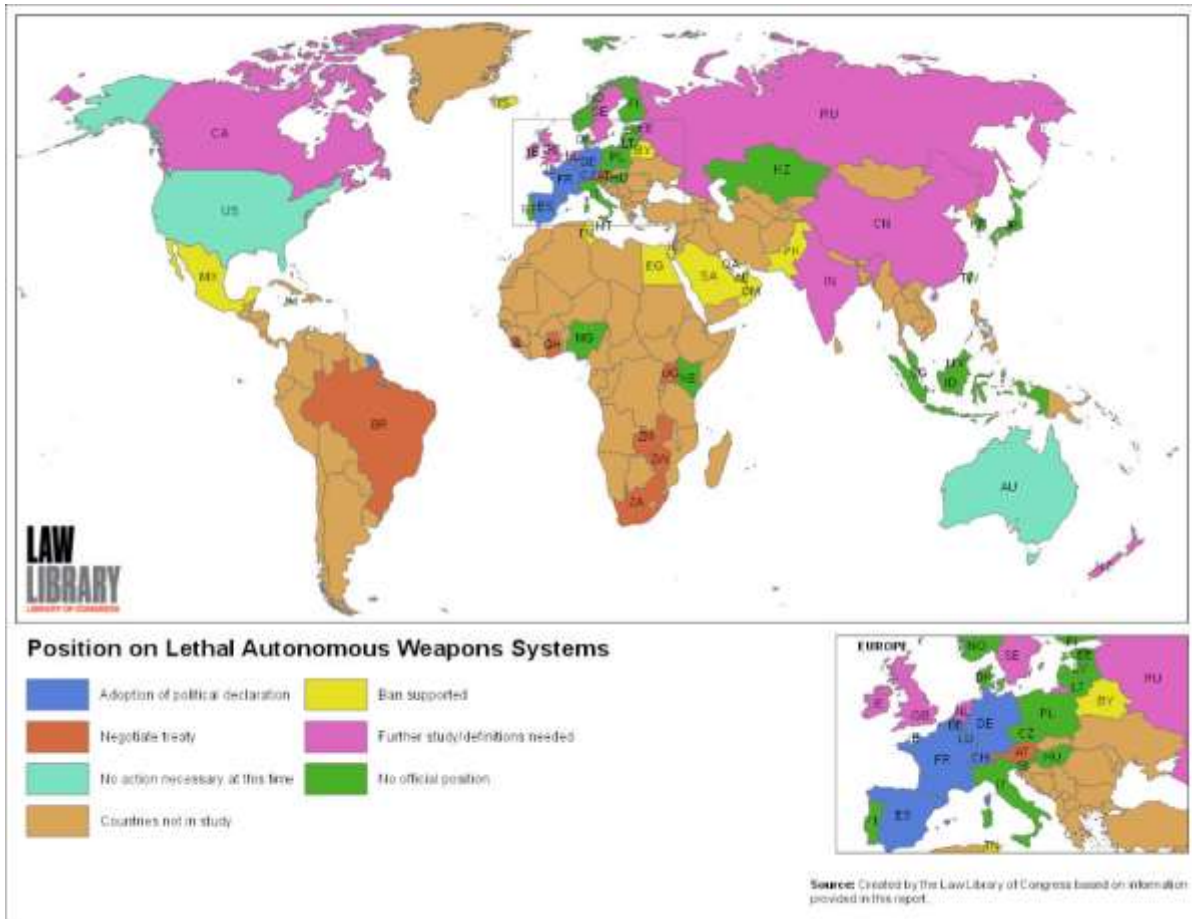


Figure 3: Regulations on autonomous weapons strategies taken from the *Regulation of Artificial Intelligence in Selected Jurisdictions* Report (The Law Library of Congress, Global Legal Research Directorate., 2019)

#### 4.1 Recommendations

Recommendations for the cybersecurity community to prepare for the introduction of AI regulations include:

- **Revise Code of Ethics.** Most respected bodies in the cybersecurity community have developed Code of Ethics to guide and prescribe what is considered to be ethical and responsible professional practice. These Codes of Ethics need to be reviewed and expanded to include the development of AI-based applications.
- **Raising Awareness.** Numerous successful campaigns have been run in the past two decades to raise awareness of cybersecurity threats and to guide cyber secure behaviour. It is now necessary to run campaigns to raise awareness in the cybersecurity community of ethical and legal considerations associated with the use and development of AI-based applications. Security awareness of the AI community is also required; security experts will need to be familiar with AI in order to adequately communicate security concerns with the AI community.
- **Compliance Tracking.** National and regional cybersecurity bodies should monitor developments in AI regulations and legislation so that cybersecurity practitioners have readily available trusted sources to consult.
- **Training.** Cybersecurity courses should include the emerging area of AI compliance in their material. Security professionals will need to broaden their knowledge to include AI.
- **Research & Collaboration.** The cybersecurity community should become involved in AI regulatory bodies, discussions and related research. Their input will be valuable to the broader AI and governing communities.
- **Increased attack surface.** Cybersecurity professionals need to keep their knowledge of attack vectors and vulnerabilities current. The increasing prevalence of AI indicates there are new attack vectors and techniques that security professionals need to be aware of.

- *Awareness of AI limitations.* Cybersecurity professionals need to be aware and understand the implications of AI limitations for automated security solutions. These may result in incorrect detection or classification. Therefore, mitigating controls need to be implemented where necessary, according to the organisation's risk posture.
- *Determine the legal consequence of security or privacy breaches due to AI.* The legal perspectives of AI failures resulting in breaches needs to be understood. For example, who is accountable: the AI product vendor or the implementing organisation?

## 5. Conclusions

AI and cybersecurity affect each other; AI can help automate cybersecurity, whereas cybersecurity is necessary to ensure trustworthy AI. Many nations have cybersecurity strategies or frameworks; however, national documents related to AI are still scarce. There is therefore a need for cybersecurity professionals to become more aware of the benefits and limitations of AI, particularly for cybersecurity applications, as well as the threat landscape and vulnerabilities related to AI in order to ensure secure implementations in their organisations.

This paper gives an overview of the development of AI regulations globally with the aim of raising awareness in the cybersecurity community of the impact these regulations will have on cybersecurity and AI. Another important development is the shift in Privacy and Data Protection legislation towards making cybersecurity a legal requirement. The paper also provides some recommendations for the cybersecurity community to become aware and prepare for the looming impact of AI regulations.

## References

- Allen-Ebrahimian, B. (2020, April 1). *Bots boost Chinese propaganda hashtags in Italy*. Retrieved from Axios: <https://www.axios.com/bots-chinese-propaganda-hashtags-italy-cf92c5a3-cdcb-4a08-b8c1-2061ca4254e2.html>
- Bocetta, S. (2020, March 10). *Has an AI Cyber Attack Happened Yet?* Retrieved from <https://www.infoq.com/articles/ai-cyber-attacks/>
- Bundesregierung. (2018). Retrieved from <https://www.ki-strategie-deutschland.de/>
- Centre for Internet Security. (2019). *CIS Controls v7.1*. <https://www.cisecurity.org/controls/cis-controls-list/>.
- Columbus, L. (2020, Dec 5). *Top 20 Predictions Of How AI Is Going To Improve Cybersecurity In 2021*. Retrieved from Forbes: <https://www.forbes.com/sites/louiscolombus/2020/12/05/top-20-predictions-of-how-ai-is-going-to-improve-cyb>
- Council of Europe. (2001). *Convention on Cybercrime*. Budapest: European Treaty Series - No. 185.
- Damiani, J. (2019). *A Voice Deepfake Was Used To Scam A CEO Out Of \$243,000*. Retrieved from Forbes: <https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=1afde0f32241>
- DarkTrace. (2021). *The Enterprise Immune System*. Retrieved from <https://www.darktrace.com/en/products/enterprise/>
- Deep Instinct. (2020). *Full protection, with a prevention first approach*. Retrieved from <https://www.deepinstinct.com/product-overview/>
- Dreyfuss, E. (2019, April 27). *2019 Security News This Week: Hackers Found a Freaky Car*. Retrieved from Wired: <https://www.wired.com/story/car-hacking-biometric-database-security-roundup/>
- Dua, S., & Du, X. (2011). *Data Mining and Machine Learning in Cybersecurity*. Boca Raton: CRC Press.
- ENISA. (2020). *AI Security Challenges: Threat Landscape for AI*. <https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges>.
- European Commission. (2020, February 19). *WHITE PAPER - On Artificial Intelligence -A European approach to excellence and trust*. [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf).
- European Parliament. (2019). *A governance framework for algorithmic accountability and transparency*. Retrieved from [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS\\_STU\(2019\)624262\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf)
- European Union. (2020, February 28). *The EU Cybersecurity Act*. Retrieved from <https://ec.europa.eu/digital-single-market/en/eu-cybersecurity-act>
- Field, M. (2017). *Graffiti on stop signs could trick driverless cars into driving dangerously*. Retrieved from The Telegraph: <https://www.telegraph.co.uk/technology/2017/08/07/graffiti-road-signs-could-trick-driverless-cars-driving-dangerously/>
- G20. (2019). *G20 AI Principles*. Retrieved from <https://www.meti.go.jp/press/2019/06/20190610010/20190610010-1.pdf>
- Gartner. (2020, February 5). *Gartner Says Over 40% of Privacy Compliance Technology Will Rely on Artificial Intelligence in the Next Three Years*. Retrieved from Gartner Newsroom: <https://www.gartner.com/en/newsroom/press-releases/2020-02-25-gartner-says-over-40-percent-of-privacy-compliance-technology-will-rely-on-artificial-intelligence-in-the-next-three-years>
- Global Commission on the Stability of Cyberspace. (2019). *Advancing Cyberstability, Final Report, November 2019*. <https://cyberstability.org/wp-content/uploads/2020/02/GCSC-Advancing-Cyberstability.pdf>.
- Global Partnership on Artificial Intelligence. (2020, June 15). *Gov.Uk*. Retrieved from Joint statement from founding members of the Global Partnership on Artificial Intelligence: <https://www.gov.uk/government/publications/joint>

- [statement-from-founding-members-of-the-global-partnership-on-artificial-intelligence/joint-statement-from-founding-members-of-the-global-partnership-on-artificial-intelligence](#)
- Heinemeyer, M. (2020). *War of the AI algorithms: the next evolution of cyber attacks*. Retrieved from <https://www.information-age.com/war-ai-algorithms-next-evolution-cyber-attacks-123491934/>
- IMB. (n.d.). *AI for cybersecurity*. Retrieved from <https://www.ibm.com/za-en/security/artificial-intelligence>
- Informations Commissioner's Office. (n.d.). *Guide to the GDPR: The Principles*. Retrieved from <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/>
- International Telecommunications Union (ITU). (2018). *Global Cybersecurity Index (GCI)*. Retrieved from <https://www.itu.int/en/ITU-D/Cybersecurity/Pages/global-cybersecurity-index.aspx>
- International Telecommunications Union. (2013). *HIPSSA – Computer Crime and Cybercrime: SADC Model Law*, <https://www.itu.int/en/ITU-D/Cybersecurity/Documents/SADC%20Model%20Law%20Cybercrime.pdf>.
- Kao, J., & Li, M. (2020, March 26). *How China Built a Twitter Propaganda Machine Then Let It Loose on Coronavirus*, 26 March, [online]. Retrieved from ProPublica: <https://www.propublica.org/article/how-china-built-a-twitter-propaganda-machine-then-let-it-loose-on-coronavirus>
- Kremlin. (2019). Retrieved from <http://www.kremlin.ru/acts/bank/44731>
- Leenen, L., & Meyer, T. (2019). Artificial Intelligence and Big Data Analytics in Support of Cyber Defence. In *Chapter 2 in Developments in Information Security and Cybernetic Wars*. IGI Global Publishers.
- Lemos, R. (2020). *Car Hacking Hits the Streets*. Retrieved from <https://www.darkreading.com/edge/theedge/car-hacking-hits-the-streets/b/d-id/1336730>
- NIST. (2013). *Improving Critical Infrastructure Cybersecurity Executive Order 13636: Preliminary Cybersecurity Framework*. <http://www.nist.gov/itl/upload/preliminary-cybersecurity-framework.pdf>.
- NIST. (2019). Retrieved from <https://www.nist.gov/document/report-plan-federal-engagement-developing-technical-standards-and-related-tools>
- Niti Aayog. (2018). Retrieved from <https://niti.gov.in/national-strategy-artificial-intelligence>
- OMB. (2020). Retrieved from <https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf>
- Organisation for Economic Co-operation and Development. (2019). *What are the OECD Principles on AI?* Retrieved from <https://www.oecd.org/going-digital/ai/principles/>
- Organisation for Economic Co-operation and Development. (2020). *Examples of AI National policies: Report for the G20 Digital Economy Task Force, Saudi Arabia 2020*. Retrieved from <https://www.mcit.gov.sa/sites/default/files/examples-of-ai-national-policies.pdf>
- Oxford Insights and the International Research Development Centre. (2020). *Government AI Readiness Index 2020*. Retrieved from <https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/5f7747f29ca3c20ecb598f7c/1601653137399/AI+Readiness+Report.pdf>
- Paris Call on Trust and Security in Cyberspace*. (2018). Retrieved from France Diplomacy.
- Participa.br*. (2020). Retrieved from <http://participa.br/estrategia-brasileira-de-inteligencia-artificial/blog/apresentacao-e-instrucoes>
- Pomares, J., & Abdala, M. (2020). The Future of AI Governance: The G20's Role and the Challenge of Moving Beyond Principles. *Global Solutions Journal*(5).
- Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2020). The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI & Soc*, <https://doi.org/10.1007/s00146-020-00992-2>.
- Schmitt, M. (2017). *Tallinn Manual 2.0: On The International Law Applicable to Cyber Operations*. Cambridge: Cambridge University Press.
- South African Government. (2020). *4IR Commission Report Recommendations gazetted*. Retrieved from SANews.gov.za: <https://www.sanews.gov.za/south-africa/4ir-commission-report-recommendations-gazetted>
- The Law Library of Congress, Global Legal Research Directorate. (2019). *Regulation of Artificial Intelligence in Selected Jurisdictions*.
- Twomey, P. (2020). *A Step to Implementing the G20 Principles on Artificial Intelligence: Ensuring Data Aggregators and AI Firms Operate in the Interests of Data Subjects*. Retrieved from G20 Insights: <https://www.g20-insights.org/authors/paul-twomey/>
- United Nations. (2015). *Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security*. A/70/174, <https://dig.watch/sites/default/files/>.
- United Nations Conference on Trade and Development (UNCTAD). (n.d.). *Cybercrime Legislation WorldWide*. Retrieved from <https://unctad.org/page/cybercrime-legislation-worldwide>
- van Niekerk, B., Ramluckan, T., & Ventre, D. (2020). Assessment of the French and Dutch Perspectives on International Law and Cyber-Operations. *19th European Conference on Cyber Warfare and Security, 25-26 June 2020*, (pp. 380-389).
- Ye, J., Cheng, X., Zhu, J., Feng, L., & Song, L. (2018). A DDoS Attack Detection Method Based on SVM in Software Defined Network. *Security and Communication Networks*, <https://www.hindawi.com/journals/scn/2018/9804061/>.
- Yerima, S., Sezer, S., & Muttik, I. (2015). High accuracy android malware detection using ensemble learning. *IET Information Security* 9(6), 313–320.



# Is Hacking Back Ever Worth it?

Antoine Lemay<sup>1</sup> and Sylvain Leblanc<sup>2</sup>

<sup>1</sup>Cyber Defence Corporation, Montréal, Canada

<sup>2</sup>Department of Electrical and Computer Engineering, Royal Military College of Canada, Kingston, Canada

[antoine.lemay@live.ca](mailto:antoine.lemay@live.ca)

[sylvain.leblanc@rmc.ca](mailto:sylvain.leblanc@rmc.ca)

DOI: 10.34190/EWS.21.020

**Abstract:** As nefarious activity in the cyber domain continues to increase, more and more actors are contemplating “hacking back” as a strategy for defence. At first glance, such deterrence may seem desirable because it intuitively offers a disincentive to the attacker to attack one’s assets; a purely defensive stance that does not cause the attacker harm may appear to do nothing to prevent or stop aggression. We must ask however, if that logic can work in practice. By looking at historical examples of cyber exchanges, we will show that many attempts to “hack back” tend to widen the scope of a conflict rather than limit it; we found that the only time when the approach works is when a credible threat of serious harm to the attacker is imposed. In fact, whether we looked at the exchanges from patriotic hackers, tit-for-tat retaliation appears to only invite additional aggression from one’s original adversary, which ultimately widens the conflict rather than appease it. On the other hand, by examining the psychological effect of the Shamoon virus, one can surmise that the United States perceived that further attacks against Iran would incur significant cost to the United States. This was likely because Iran’s capabilities to reverse engineer the attacks which would leave the United States vulnerable in the context of an asymmetric Iranian response, a counter-value proposition caused by common and increasing automation in the management of United States critical infrastructure. Similarly, even though United States retaliation for the Sony Picture hack was proportional in its effect, it demonstrated the ability to significantly damage North Korea’s core connectivity, a fact that can be interpreted as signalling a crippling counter-force capability. In both instances, it seems that it was the threat of escalation, rather than retaliation itself, which proved effective. These observations lead to further questions which will be explored in the paper. Notably, whether cyber threat projection settles in escalation ladders frameworks as was the case with nuclear weapons, or whether the need for secrecy, required to maintain capability, interferes with the need for establishing credibility in threats of escalation.

**Keywords:** cyber retaliation, cyber deterrence, cyber conflict, cyber escalation

---

## 1. Introduction

All indications are that malicious activity in the cyber domain is here to stay. The news is replete with the reporting of new cyber-attacks, and the list of victims is ever increasing. As the well-known public-interest technologist and cyber-security expert Bruce Schneider is fond of saying, “attacks always get better; they never get worse” (Schneider, 2015). Furthermore, as advanced economies become increasingly reliant on digital technologies, a trend exacerbated by the current pandemic, it would be reasonable to expect that cyber-attacks can become more prevalent and cause more damages. The cyber realm is thus an extremely attractive avenue for attackers to exploit. It provides unprecedented access for regional players, such as North Korea, who can credibly threaten assets in the continental United States. More importantly, cyber-attacks come with a certain amount of impunity. This is in part because of the perceived anonymity of the cyber realm; many attackers imagine they can cause mayhem without suffering consequences because of the perceived difficulties that victims have in making public attribution. The matter is further complicated by the legal regime in the cyber realm where, even when the identity of attackers is known, it is often illegal to impose a sanction on them or impractical to do so when the legislative framework allows it. For example, malicious actors might be in a remote jurisdiction that is sympathetic to the attackers or the cyber-attack might not be severe enough to rise to meet the threshold required to justify retaliation in the kinetic domain. This ability to project power without consequence is obviously extremely attractive to certain classes of actors, creating a potential asymmetry between small attackers and large defenders.

In order to impose a cost on attackers, many have suggested that there may be benefit in “hacking back”, i.e. taking a much stronger approach than simply defending oneself by performing offensive cyber operations on the perpetrators of these attacks. Discussions about such a potential “hack back” strategy have been increasing in the private sector since the wave of cyber espionage surrounding the development of the Lockheed Martin F-35 Lightning II fighter jet (Kallberg, 2015). The concept has also gained prominence due to the actions of the United States Cyber Command (USCYBERCOM) targeting the Russian Internet Research Agency (IRA) prior to the

2018 midterm elections (Greenberg, 2019). In theory, this strategy appears to make perfect sense. By imposing a cost on the attackers, defenders render the option of the cyber-attack much less attractive. In fact, game theory would suggest that if the cost is high enough, a defender might even deter the opposing side from attacking entirely. One important question remains however: does it work?

This paper aims to look at historic examples of retaliatory threats in the cyber realm to ascertain if retaliatory threats are effective at preventing further cyber harm. The paper will start by providing background information on hacking back (i.e., retaliatory offensive cyber operations) and on the literature describing it. Then, the paper will then examine a certain number of historic examples of cyber conflict that included retaliation to determine if retaliation played a role in deterrence. This will lead into an analysis of the salient points which will derive general concepts from the historic examples examined. Finally, the paper will summarize its conclusions.

## **2. Background**

The idea of taking retaliatory action against an adversary that has carried out a cyber-attack by hacking back has been around for quite some time. Examples from the scientific press describe two early efforts, where the United States Pentagon was said to have used offensive agents to disrupt the Electronic Disruption Theater, a so-called hactivist organization, which was carrying out a denial of service attack in 1988 (Jayaswal et al., 2002). The same source also described an early effort by the World Trade Organization where the WTO server host redirected the emails that were targeting the organization to the hactivist group that encouraged citizens to send spurious emails to the WTO.

Many problems are associated with the idea of hacking back, as discussed in (Jayaswal et al., 2002). First is the idea of incorrect identification. The difficulty of attribution in cyber space is a very topical problem which is the subject of current research, introducing a variety of approaches such as a Detection Maturity Level Model and a Cyber Threat Intelligence Model (Mavroeidis and Bromander, 2017). The study of these models is beyond the scope of this paper, but it is evident that difficulties in attribution make the risks associated with hacking back high, and thus representing a significant risk to organizations that are considering it.

Second is the liability to which an organization exposes itself by hacking back which are covered in a variety of sources such as the excellent note from the Columbia Journal of Transnational Law (Messerschmidt, 2013). There is an obligation to prevent transboundary harm from cyber-attacks, but the issues of territorial sovereignty in the cyber realm are complicated and lack a firm legal framework that allows for the enforcement of international norms.

Nonetheless, there has been increasing pressure to allow for hacking back in the popular press (Matthews, 2013). We will now examine some historical examples of such cyber retaliation.

## **3. Historic examples**

One the first documented cases of so-called “cyberwar”, we can observe many groups of patriotic hackers going after opposing nations after an international conflict. Notably, we can cite the Israel-Palestine “cyberwar” (Allen and Demchak, 2003) and the conflict surrounding the downing of a United States spy plane near China (“Chinese and American hackers declare ‘cyberwar,’” 2001). Both cases saw periods of heightened international tension between two protagonists leading to cyber activists taking aggressive actions against cyber assets in the other protagonist’s jurisdiction, which is in turn met with retaliation. In both cases, the retaliation invited more and broader retaliation against the opposing side; rather than be feeling enough pain to stop attacks, adversaries widened with the conflict. The situation even expanding “horizontally” to include targets in friendly countries in the case of the Israel-Palestine conflict (Allen and Demchak, 2003) and reaching hundreds of affected sites in the spy plane conflict (Tank, 2001).

This type of retaliatory action clearly has no effect on deterrence. In fact, the Israel-Palestine cyber conflict is still ongoing and with a significant expansion in its scope, with Israel using pre-emptive kinetic strikes against Palestinian hackers (Hay Newman, 2019). From this, one could theorize that what is present here is tit-for-tat retaliation of increasing severity, which will continue to increase until the international tensions are resolved by other means. If the conflict is not somehow resolved, it is difficult to imagine what would cause the retaliation to do anything but to continue to escalate. Notably, Israel-Palestine conflict saw two distinct kinds of retaliation,



a quick horizontal retaliation increasing the number of targets involved in conflict was observed, along with a vertical escalation by going from cyber to kinetic actions.

Another case is the current Israel-Iran cyber exchanges in which Iran is targeting Israel water infrastructure (Warrick and Nakashima, 2020) and where Israel launched retaliatory cyber strikes on an Iranian port (Bergman and Halbfinger, 2020). According to Halbfinger, this represents a case where Israel is using a proportional cyber response to signal Iran to stop attacking Israel cyber infrastructure. However, in its public declarations, Israel's cyber czar was warning that a "cyber winter" was forthcoming (Bob, 2020) because of the cyber-attacks from Iran. This indicates that the retaliation is expected to escalate, rather than to provide deterrence. In fact, the deterrence impact is so limited that some authors argue that the retaliatory strike might be viewed as an pretext to disrupt Iran's nuclear supply chain instead (Work and Harknett, 2020).

One thing of interest in this exchange is that both sides take great lengths to spin the damage as minimal. Israel has underlined the fact that the initial intrusion in the water system was stopped and Iran's port authority has minimized the extent of the disruption at the port. This rationalization of the cost indicates that the intended deterrent effect, meaning imposing a cost sufficiently large for the attacker that it would outweigh the benefits, is unlikely to play a significant effect in future decisions. In fact, since both sides can declare victory, it seems likely to invite future exchanges rather than limit them. This would also lend credence to the escalatory expectations of the Israeli cyber czar.

Another case involving Iran warrants examination. In contrast, this conflict involves the United States rather than Israel. Following the Stuxnet attack, Iran launched the Shamoon attack which rendered approximately 30,000 computers useless at Saudi Aramco (Mahdi, 2012) and which also crippled office computers at RasGas, a Quatary gas company (Zetter, 2012). While the retaliatory strike was not aimed directly against United States assets, the Americans perceived Shamoon as a precursor of future attacks. The fact that the virus showed a picture of a burning American flag also strongly hinted at the intended ultimate target of the attack. In classified documents leaked by Snowden, the NSA reveals that components of Shamoon were inspired by western cyber-attacks on Iran infrastructure and that the Iranians were learning from cyber-attacks against them (Zetter, 2015). Based on these observations, one can conclude that the Shamoon virus has successfully altered the strategic calculus of the United States by imposing a cost to future American actions. Specifically, that future cyber action would empower the Iranian to perform increasingly devastating retaliations by learning from the attacks to which they were subjected. So, while we cannot objectively assert that the United States did not perform any offensive cyber operations against Iran, the fact that there is evidence that the calculus was changed allows us to say that this Iran was successful in using cyber-attack as a deterrent against future American retaliation.

Although none were made public, the United States may have performed additional cyber operations against Iran. Even if that had been the case, the United States would have needed to pause and consider the cost-benefit ratio of such operations, now having to consider the risk of knowledge transfer to this dangerous adversary. This is a significant change from a starting position of considering the attack as having no costs, a fact that would likely lead to a significant reduction in the range of cyber actions considered. From Iran's point of view, this should still be considered a success in the face of a dramatic power imbalance where one side had perceived immunity due to overwhelming force.

What is most interesting in this case is that the retaliation was not targeted at the state that Iran hoped to deter, but rather one of its allies. The United States did however understand the signal that future attack would be costly as, Iran would learn from attacks and could use the knowledge gained against the instigators of retaliatory attacks to great effect. The fact no assets were targeted in the United States limited the potential for tit-for-tat retaliation. Also, the cost is bounded by the U.S. imagination rather than by actual results on the ground. This prevents the rationalization that the costs were not so bad based on observed results.

Our last historical example involves the United States and North Korea. Around the time of the release of the movie "The Interview", in which a TV personality is tasked by the CIA to assassinate the North Korean leader, hackers linked to North Korea attacked the Sony Pictures movie studio. In response, the United States administration promised a proportional retaliation and an Internet outage identified as part of the response was observed in North Korea (Woldt, 2015; Wroughton and Rajagopalan, 2014); interestingly, the response also spilled over in the diplomatic realm as it included new sanctions. In this case, while the retaliation did not stop North Korean cyber-attacks in general, no further attacks aimed at restricting freedom of expression were

observed, which was considered “a bright red line” in official statements by the United States administration (Bennett, 2015).

In this example, we see the only case where the retaliation has a counter-force effect (directly affecting adversary abilities) rather than a purely counter-value effect (by influencing the adversary’s behaviour). It is worthwhile to note that civilian Internet penetration in North Korea is extremely limited and that it is subject to strict control. Causing an Internet outage therefore impacted mostly the government and military forces. Of note, such an outage limits the ability of North Korean cyber forces to perform their normal operations by projecting power beyond their national borders, including the accumulation of foreign currency for the regime. As such, the denial of Internet access can have a disproportionate effect designed to deter North Korea from going over the American bright red line associated with freedom of expression. By avoiding specific actions targeting American freedom of expression, the North Korean regime can continue to perform offensive cyber operations on which they rely to bypass sanctions (Murdock, 2020). This creates a reasonable means for de-escalation of retaliatory cyber actions.

#### **4. Analysis**

In the examples discussed above, we have observed some instances of successful deterrence via retaliation and other instances where retaliation had escalatory effects. In that sense, there does not seem to be a general principle through which the ability and willingness to strike back deters attacks. That being said, we cannot completely dismiss the deterrence effects of hacking back.

In an inaugural lecture at King’s College in London by the head of the U.K. National Cyber Security Center (Martin, 2020), Martin says:

*“The second reason for possessing, and being willing to use, cyber capabilities which is often discussed, is about deterring cyber-attacks. Here the evidence is decidedly mixed. By that I don’t mean it’s inconclusive: I mean in some areas it clearly works and in others it clearly doesn’t.*

*It seems to work in respect of the second level of activity, the direct degradation of adversary infrastructure. This is simple deterrence not so much by denial but by destruction: we have located the attack infrastructure and destroyed it so it cannot be used. UNITED STATES Cyber Command’s operations, and those of allies, in this area seem to be achieving significant effect. If it is true, as some American media reports have claimed, that Western cyber operatives are looking at ways to take out the infrastructure used by the organised criminals responsible for the ransomware attacks that have included hospitals and other healthcare providers, I will be the first to applaud.*

*Where cyber-attacks don’t work, in my strong view, is as a psychological deterrent to attackers. I don’t say this as a matter of philosophical conviction; I would love it to be true that cyber-retaliation deters attackers.*

*But it’s not true.*

*In all my operational experience, I saw absolutely nothing to suggest that the existence of Western cyber capabilities, or our willingness to use them, deters attackers. Nor have I seen any convincing research.”*

So, an initial hypothesis would be that counter-value operations aimed at producing a psychological effect should be eschewed in favour of counter-force actions. Unfortunately, this runs counter to the facts as we have previously presented them. First, the most demonstrably successful use of deterrence is the Shamoon virus which is a form of counter-value attack. Second, we must also say that results of attempts to degrade an adversary’s cyber-attack infrastructure are still mixed. While the operation to dismantle the Islamic State’s network infrastructure was deemed successful (Marks, 2020), use of a similar strategy against the Russian IRA in the leadup to the 2018 midterm elections was received with more scepticism (Greenberg, 2019). In fact, a number of the same actors, along with some new players, have been spotted attempting to interfere with the 2020 United States’ election (O’Neil, 2020).

An alternative hypothesis would be that capabilities have not been deployed effectively as a threat; their presence has not been sufficient to provide deterrence. To give an example, there is no doubt that cyber offensive capabilities of the Five Eyes intelligence alliance (Australia, Canada, New Zealand, United Kingdom, United States of America) are significant. It can be argued however, that they have shown a propensity to either

not retaliate or to only engaged in limited and proportional retaliation. Furthermore, without a full accounting of the cyber defensive capabilities of the Five Eyes intelligence alliance, an attacker might believe that they can carry out cyber-attack while remaining undetected or without having their actions attributed to them. If these premises hold, one could presume that it would not be possible to retaliate against the perpetrator, even if the offensive capability to do so existed. As such, an adversary might still *believe*, wrongfully or not, that they can act with impunity or with only the risk of suffering minimal consequences.

One could argue that the *perception* of potential consequences is more important than their actual calculus; it may well be the case that poor signalling is the main hurdle facing retaliation-based deterrence. In the North Korean example we discussed (Bennett, 2015), the United States demonstrated the kind of pain it was willing to inflict as well as its ability to do it. Similarly, the demonstration of capability from Shamoon (Mahdi, 2012) sent a clear and concise message that attacks would be repurposed. In that example, the United States might even have perceived the costs of future attacks to be even greater than what would be expected of Iranian capabilities. On the other hand, the Israel-Iran series of attacks (Warrick and Nakashima, 2020) sent muddled messages. It is not entirely clear the two events are linked, even if allusions are made that they are. Was the attack on the port the extent of the Israeli capability or just a measured response that could be escalated even further? Was the attack just an attack of opportunity to disrupt the nuclear program? How does the value of the delay in shipping compare to the propaganda value gained from obtaining access to the Israeli water supply? Having the adversary arrive at the wrong conclusion in any of these questions might lead them to a calculus that would be detrimental to the party doing the signalling.

Given that a major problem with the hacking back approach is that the aggressor may incorrectly perceive that the attack cannot be attributed to them, for example by tunnelling the attack through a 3<sup>rd</sup> country or by operating by funding third parties, the continued classification of attacks and poor attribution capabilities will continue to be problematic. Operational cyber capabilities, whether on offense or defence, often are the subject of the greatest secrecy because of the desire to protect one's sources and methods. For example, it is widely believed that attribution in the cyber realm is difficult (Mavroeidis and Bromander, 2017). This difficulty in attribution may not be as pronounced as an actor can supplement their technical attribution efforts with other means such as human intelligence. In such a case, the attacker may erroneously believe that it is impossible to trace the attack back to them because they can obfuscate technical indicators, while complete attribution may be possible through other means. Similarly, defenders possessing technical detection measures that could employ technical intelligence to attribute attacks to specific actors with great confidence, could not deter an adversary from attack unless they revealed their detection capabilities. This is in turn problematic because defenders know that if their adversaries know about their detection capabilities, they may be able to devise a way to exploit the detection capabilities by indicators carrying out counter-surveillance operations (Knight and Leblanc, 2009). There is no ready end in sight to this particular conundrum.

We must also consider an alternative scenario in which it is advantageous for the adversary to de-escalate. If the pain caused by the adversary's retaliation is less than the pain one's attack causes the adversary, the threat of retaliation is not credible. In the patriotic hackers' example (Allen and Demchak, 2003), one could well imagine that the pain of a few more web defacements is much lower than the pain of admitting the other side won. At the other end of the spectrum, allowing North Korea to return to its normal operations instead of facing a massive escalation of the conflict (Murdock, 2020), presents a much clearer incentive for de-escalation.

When examining all of these facts, we can conclude that the main obstacle to deterrence may be the lack of credibility of the cyber retaliation threat. This could be either a problem of perception on the part of the aggressor (e.g., they think they will not get caught or they believe that their victim will not have the will to retaliate) or it could be a lack of sufficient demonstrated incentives to alter behaviour. In that sense, retaliatory attacks might only be effective to influence behaviour in the face of disproportionate threats, in particular threats of unbounded escalation, and with frequent demonstration of capabilities. Such public demonstration would ensure adversary correctly interpret both the likelihood of retaliation and correctly perceive that the costs outweigh the gains.

This line of thinking represents intellectual ground that has been well tread in the realm of nuclear strategy. Deterrence based on retaliatory threats is the underpinning of the escalation ladders going to mutually assured destruction. In the absence of a well-publicised crippling counter-force capability, the adversary may always threaten to escalate the conflict to deter retaliation, until no escalation is possible. In fact, the Nuclear Arms

Control Association is already worried about the public commitment to nuclear retaliation against certain cyber-attacks (Klare, 2019). These unbounded escalation scenarios lead to brinksmanship, which in turn create many opportunities for erroneous calculus and catastrophe.

## 5. Conclusion

There is good reason to be sceptical of retaliatory offensive cyber operations or hack back activities. In many cases, they can invite further retaliation and thereby escalate the conflict. If the aggressor perceives that they can defend against or hide from the retaliation, or if they believe that they can reply to it in-kind, there is no incentive to back down. To be successful in deterring future aggression, hack back actions need to be paired with a credible threat. This means that the aggressor calculus needs to be altered so that they perceive that they must change course when presented with suitable incentive to do so. In that sense, increased signalling should be part of offensive cyber operations, not only to establish the willingness to retaliate, but also to threaten increased future harm.

Unfortunately, there are limits to the application of these requirements for success. First, the secrecy surrounding cyber capabilities, both offensive and defensive, negatively impacts the adversary's perception of one's own capabilities, and thereby increasing the adversary's propensity to forego retaliation and to enter in a tit-for-tat. Also, the need to send strong retaliatory messages may lead to unbounded escalation and brinksmanship, which is known to be undesirable from the history of nuclear strategy.

It is perhaps this similarity to nuclear escalation that may yield interesting future research. A large body of work regarding the credibility of threats and the use of signalling was developed to deal with nuclear weapons. Investigating how these efforts can be transferred to the cyber realm might provide insight to avoid the escalation traps observed with nuclear weapons. Finally, a more in-depth investigation with the benefit of hindsight of the recent by USCYBERCOM to focus more on pre-emptive counterforce action than counter-value retaliation would be interesting.

## References

- Allen, P., Demchak, C., 2003. The Palestinian-Israeli Cyberwar. *Mil. Rev.* 83, 52.
- Bennett, C., 2015. NSA chief: Sony hack a "red line" for US [WWW Document]. *TheHill*. URL <https://thehill.com/policy/cybersecurity/229005-nsa-chief-us-had-to-draw-red-line-after-sony-hack> (accessed 2.8.21).
- Bergman, R., Halbfinger, D.M., 2020. Israel Hack of Iran Port Is Latest Salvo in Exchange of Cyberattacks. *N. Y. Times*.
- Bob, Y.J., 2020. Israeli cyber czar warns of more attacks from Iran [WWW Document]. *Jerus. Post JPostcom*. URL <https://www.jpost.com/israel-news/israeli-cyber-czar-warns-of-more-attacks-from-iran-629577> (accessed 2.8.21).
- Chinese and American hackers declare "cyberwar" [WWW Document], 2001. *the Guardian*. URL <http://www.theguardian.com/technology/2001/may/04/china.internationalnews> (accessed 2.8.21).
- Greenberg, A., 2019. US Hackers' Strike on Russian Trolls Sends a Message—but What Kind? *Wired*.
- Hay Newman, L., 2019. What Israel's Strike on Hamas Hackers Means For Cyberwar. *Wired*.
- Jayaswal, V., Yurcik, W., Doss, D., 2002. Internet hack back: counter attacks as self-defense or vigilantism? In: *IEEE 2002 International Symposium on Technology and Society (ISTAS'02). Social Implications of Information and Communication Technology. Proceedings (Cat. No.02CH37293)*. Presented at the IEEE 2002 International Symposium on Technology and Society (ISTAS'02). *Social Implications of Information and Communication Technology. Proceedings (Cat. No.02CH37293)*, pp. 380–386.
- Kallberg, J., 2015. A Right to Cybercounter Strikes: The Risks of Legalizing Hack Backs. *IT Prof.* 17, 30–35.
- Klare, M.T., 2019. Cyber Battles, Nuclear Outcomes? Dangerous New Pathways to Escalation | Arms Control Association [WWW Document]. URL <https://www.armscontrol.org/act/2019-11/features/cyber-battles-nuclear-outcomes-dangerous-new-pathways-escalation> (accessed 2.8.21).
- Knight, S., Leblanc, S., 2009. Chapter 16: When Not to Pull the Plug – The Need for Network Counter-Surveillance Operations. In: *Czosseck, C., Geers, K. (Eds.), The Virtual Battlefield: Perspectives on Cyber Warfare, Cryptology and Information Security Series*. pp. 226–237.
- Mahdi, W., 2012. Saudi Arabia Says Aramco Cyberattack Came From Foreign States. *Bloomberg.com*.
- Marks, J., 2020. Analysis | The Cybersecurity 202: Here's the inside story of Cyber Command's campaign to hack ISIS. *Wash. Post*.
- Martin, C., 2020. Ciaran Martin: "Cyber weapons are called viruses for a reason: statecraft, security and safety in the digital age." [WWW Document]. *Strand Group*. URL <https://thestrandgroup.kcl.ac.uk/event/ciaran-martin-cyber-weapons-are-called-viruses-for-a-reason-statecraft-security-and-safety-in-the-digital-age/> (accessed 2.8.21).
- Matthews, C.M., 2013. Support Grows to Let Cybertheft Victims "Hack Back." *Wall Str. J.*

- Mavroeidis, V., Bromander, S., 2017. Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence. In: 2017 European Intelligence and Security Informatics Conference (EISIC). Presented at the 2017 European Intelligence and Security Informatics Conference (EISIC), pp. 91–98.
- Messerschmidt, J.E., 2013. Hackback: Permitting Retaliatory Hacking by Non-State Actors as Proportionate Countermeasures to Transboundary Cyberharm Note. *Columbia J. Transnatl. Law* 52, 275–324.
- Murdock, J., 2020. North Korea internet use spikes as regime relies on hacking and cryptocurrencies to circumvent sanctions [WWW Document]. *Newsweek*. URL <https://www.newsweek.com/north-korea-internet-use-surges-cybercrime-cryptocurrency-evade-sanctions-recorded-future-1486637> (accessed 2.8.21).
- O’Neil, P.H., 2020. The Russian hackers who interfered in 2016 were spotted targeting the 2020 US election [WWW Document]. *MIT Technol. Rev.* URL <https://www.technologyreview.com/2020/09/10/1008297/the-russian-hackers-who-interfered-in-2016-were-spotted-targeting-the-2020-us-election/> (accessed 2.8.21).
- Schneider, B., 2015. SHA-1 Freestart Collision - Schneier on Security [WWW Document]. URL [https://www.schneier.com/blog/archives/2015/10/sha-1\\_freestart.html](https://www.schneier.com/blog/archives/2015/10/sha-1_freestart.html) (accessed 2.8.21).
- Tank, R., 2001. CNN.com - China-U.S. cyber war escalates - May 1, 2001 [WWW Document]. URL <https://www.cnn.com/2001/WORLD/asiapcf/east/04/27/china.hackers/index.html> (accessed 2.8.21).
- Warrick, J., Nakashima, E., 2020. Foreign intelligence officials say attempted cyberattack on Israeli water utilities linked to Iran. *Wash. Post*.
- Woldt, T., 2015. North Korea Web outage was retaliation for Sony hack, lawmaker says [WWW Document]. *Dallas News*. URL <https://www.dallasnews.com/business/technology/2015/03/17/north-korea-web-outage-was-retaliation-for-sony-hack-lawmaker-says/> (accessed 2.8.21).
- Work, J., Harknett, R., 2020. Troubled vision: Understanding recent Israeli–Iranian offensive cyber exchanges. *Atl. Council*.
- Wroughton, L., Rajagopalan, M., 2014. Internet outage seen in North Korea amid U.S. hacking dispute. *Reuters*.
- Zetter, K., 2012. Qatari Gas Company Hit With Virus in Wave of Attacks on Energy Companies. *Wired*.
- Zetter, K., 2015. The NSA Acknowledges What We All Feared: Iran Learns From US Cyberattacks. *Wired*.

# EU Digital Sovereignty: A Regulatory Power Searching for its Strategic Autonomy in the Digital Domain

Andrew Liaropoulos

University of Piraeus, School of Economics, Business and International Studies,

Department of International and European Studies, Piraeus, Greece

Laboratory of Intelligence and Cyber-Security

[aliarop@unipi.gr](mailto:aliarop@unipi.gr)

[andrewliaropoulos@gmail.com](mailto:andrewliaropoulos@gmail.com)

DOI: 10.34190/EWS.21.037

**Abstract:** Digital technologies have gradually affected the way societies interact, how companies deliver services and how people are governed. Policymakers around the world have realized the importance of digital technologies on their countries' security and autonomy and have issued sovereignty claims regarding cyberspace. The European Union - an actor that aims to ensure that governments, the private sector, civil society organisations and end users around the world promote an open, free, and secure cyberspace - has recently added the concept of digital sovereignty in its political vocabulary. Taking for granted that there is no widely accepted and comprehensive approach regarding digital sovereignty, this paper will analyse the European discourse on digital sovereignty. It will first review the ambiguous concept of sovereignty and then explore the way it can be applied in the European digital domain. The aim is to highlight the dilemmas and constraints that the EU is facing in relation to regulating the digital domain, avoiding technological protectionism, promoting cyber-resilience, and understanding the game of digital geopolitics.

**Keywords:** EU, sovereignty, digital sovereignty, digital autonomy

---

## 1. Introduction

Over the past years, the EU has faced numerous security challenges in the digital realm (Carrapico & Barrinha 2017; Christou 2019) and gradually developed various policies and institutions (Kapsokoli 2020) that aim to safeguard its citizens in the digital domain. It is in this context that the concept of digital sovereignty (Timmers 2019; Christakis 2020; Hobbs 2020) has become the latest buzzword in the corridors of Brussels. Bearing in mind that the concept of sovereignty is regarded as a foundational concept of international politics, even though an essentially contested one (Biersteker & Weber 1996; Krasner 1999), it is challenging to examine how this concept is applied in the digital realm. Cyber sovereignty, internet sovereignty, information sovereignty, technological sovereignty and data sovereignty are only some of the terms that have entered the relevant literature over the last two decades (Liaropoulos 2017; Mueller 2020) and enriched the debate on the nature of sovereignty. It is even more thought-provoking to consider how digital sovereignty is understood in the context of the EU, since the term sovereignty does not even appear in the EU constitutive treaties. In view of the fact, that the very idea of 'European sovereignty' (Leonard & Shapiro 2019) - let alone a digital one - is a rather ambiguous concept, it is necessary to develop an understanding of this hard to define term.

To do that, the paper will first analyse the concept of sovereignty. The latter embodies an internal and an external/international dimension. The first refers to the idea of a supreme decision-making and enforcement authority over a given territory and population. The latter refers to the absence of a supreme international authority and therefore the independence of sovereign states. This typology will also be applied in the case of digital sovereignty. Thus, the paper will review EU's capacity in regulating its digital domain, but also in acting autonomously in the arena of digital geopolitics. Considering that digital technologies are profoundly affecting every aspect of our societies, from cybersecurity and techno-nationalism to data localization and cyberspace governance, it is important to study how Europe aspires to tackle these issues.

## 2. From sovereignty to cyberspace sovereignty

There is no doubt, that sovereignty is one of the most controversial terms in international politics. To put it in plain words, sovereignty is about a power that has no higher power above it. Despite its rich intellectual history, there is still disagreement regarding the role and significance of sovereignty in contemporary politics. From the early works of Jean Bodin, Thomas Hobbes, and Immanuel Kant to contemporary scholars like Carl Schmitt and Stephen Krasner, state sovereignty has been related to the people's right to establish an identity and protect

self-determination against external interference, but also to domestic atrocities and genocide (Slomp 2008, 33). Sovereignty is closely related to concepts like governance, security, independence, and democracy.

A popular perception of sovereignty is the one that discerns between the internal and the external/international sovereignty. The former is understood as the supreme power that the state has over its citizens within its borders and therefore the supreme decision-making and enforcement authority over a specific territory and towards a population. The latter is understood as the absence of a superior power to states. International sovereignty represents the principle of self-determination in the absence of a supreme international authority (Slomp 2008, 40-42). In the international context, sovereignty meets independence. States, regardless of their power status, are considered sovereign and independent and enjoy equal rights under international law (Christakis 2020, 5). This is also reflected in the UN Charter, in the principles of sovereign equality, Article 2(1), territorial integrity Article 2(4) and non-intervention Article 2(7) (Aalberts 2016, 186). Therefore, sovereignty does not only signify authority within a distinct territorial entity, but it also implies equal membership of the modern states system.

Sovereignty - a concept that continues to trigger conceptual battles - becomes even more complex when applied in a domain with no clear physical lines. To begin with, is cyberspace beyond the reach of state sovereignty? How do states perceive and exercise their sovereignty in relation to information and communication technologies - ICT? Is it possible for states to apply territorial jurisdiction in a borderless space? Can cyberspace be governed and regulated, or should we perceive it as a case of global commons? (Wu 1997; Mueller 2010; DeNardis 2014). Discussing the idea of sovereignty in relation to cyberspace has a short, but nevertheless, rich history.

In the early days of the Internet development, the notions of territory and governance seemed rather irrelevant in this human-made and spaceless domain. Indicative of this are the views of John Perry Barlow, the founder of the Electronic Frontier Foundation (EFF). In 1996, in his manifesto titled *A Declaration of the Independence of Cyberspace* he states the following: "Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather. We have no elected government, nor are we likely to have one...Cyberspace does not lie within your borders." (1996). For utopians, cyberspace is a separate entity with little or ideally no top-down regulation since it can be self-regulated. Again, in the words of Barlow "Where there are real conflicts, where there are wrongs, we will identify them and address them by our means. We are forming our own Social Contract. This governance will arise according to the conditions of our world, not yours. Our world is different." (1996).

From a different theoretical angle, Mueller argues against sovereignty in cyberspace. His thesis is that we cannot impose jurisdictional borders on cyberspace, simply because borders in cyberspace do not align with the territorial jurisdictions of states. According to Mueller, cyberspace due to its unique technical structure, should be approached as global commons and regulated as such (2020).

The issue of exercising sovereignty in relation to cyberspace comes down to whether the latter is indeed a non-territorial and borderless domain. To approach this issue, it is critical to distinguish between the physical and non-physical elements of cyberspace. Stephen Gourley differentiates between the domain (the medium) and the space. He refers to the physical aspects as the cyber domain - anything that enables users to transmit, store, and modify digital data - and to the non-physical aspects as cyber space (written as two words). Cyber activities take place through the cyber domain in cyber space (Gourley 2013, 278). The cyber domain is an artificial and human-made construct with geographical ties over a specific territory. This infrastructure is terrestrially based and, therefore, not immune from state sovereignty. Thereby, states can exercise their sovereignty through the cyber domain. According to Gourley the territoriality principle allows states to control cyber activities occurring within and across their borders, and the effects principle gives them jurisdiction over external activities that cause effects internally. Applying the same analogy for the non-physical aspects of cyberspace is tricky. The reason is that there is no universal approach regarding data/information sovereignty (Gourley 2013, 279-280).

The exercise of state sovereignty in cyberspace raises two issues (Cornish 2015, 157). First, it conflicts with the idea of cyberspace as a global common; and second, it could result in the fragmentation of cyberspace. Regarding the idea of cyberspace as a global common, in sharp contrast to sea and air that are limited by geographical boundaries, cyberspace is a human-made domain that is not constrained by physical space. In relation to size, cyberspace is unbounded. In sharp contrast to the domains of land, sea and air that are limited,

cyberspace itself is growing and evolving as information technology expands and develops. Contrary to popular belief, cyberspace does not meet the legal criteria of global commons and is not free as are the atmosphere and ocean (Betz & Stevens 2011, 107). Paradoxically, although cyberspace seems borderless, it is bounded by the physical infrastructures that facilitate the transfer of data and information. Such infrastructures are mostly owned by the private sector and are located within the sovereign territory of states. Therefore, it would be more precise to argue that cyberspace comprises a global common infrastructure, but is not a global common (Cornish 2015, 158).

Regardless of their theoretical departure, scholars that advocate the concept of cyberspace sovereignty, argue that the latter is a useful concept to deal with cybersecurity issues, develop the necessary norms of responsible state behavior and gradually construct an international cyber-order (Mueller 2020, 784). In 2015, the United Nations Governmental Group of Experts confirmed that sovereignty, international norms, and principles that derive from state sovereignty apply to state conduct of ICT activities and to their jurisdiction over ICT infrastructure within their territory (UNGA 2015). Thus, the principle of sovereignty in relation to cyberspace is essentially uncontested. What is contested though, are the norms that relate to sovereignty, since states do not share the same understanding on the free flow of information, privacy, anonymity and national (cyber)security. Over the past years, many states, mainly authoritarian ones, have used the sovereignty card - information, data, technological and digital - to restrain internet freedoms and exercise digital surveillance (Kamasa 2020). Having conceptualized an understanding of what sovereignty entails in relation to cyberspace, we will turn our analysis to the concept of digital sovereignty in the context of the EU.

### **3. The European digital sovereignty: the power to regulate**

The EU is a unique actor. It is not a state, but the Union is much more than just an intergovernmental organization. Putting for a moment aside the theories of European integration, the democratic deficit and the legal nature of this sui generis organization, any discussion about European sovereignty, and especially a digital one, raises eyebrows. After all, the idea of removing sovereignty from national capitals to recreate it in Brussels is a complex issue, in both conceptual and political terms. Whether European sovereignty implies the weakening of national sovereignties, the creation of a shared sovereignty or the construction of collective sovereignty is a strong theoretical exercise that meets the political reality of European power politics.

Christakis correctly points out, that in strictly legal terms, digital sovereignty does not apply to the traditional understanding of sovereignty that was analysed above, since regardless of their economic power, technological giants and digital platforms like Google, Amazon, Facebook, Apple, Microsoft - the so-called GAFAM - and others, are in no position to exercise any legislative or jurisdictional authority on the EU member-states (2020, 5-6). This forces us to approach the very concept of digital sovereignty in different terms.

A quick review of the policy papers that the EU has published and the statements that key EU officials have made over the last three years, demonstrates that the digital sovereignty debate, highlights the anxiety that the EU will not be able to effectively regulate its digital universe, protect its citizens' data and compete successfully with China and the US in the arena of digital geopolitics (European Commission 2019; Christakis 2020; Hobbs 2020). References made by EU policymakers to digital or technological sovereignty and digital autonomy imply that if the EU is a weak actor in the digital domain, this will affect its ability to regulate its digital services, protect its infrastructure and values and be able to shape the development of global norms regarding cyberspace governance.

The quest for digital sovereignty is rooted in a perception that the EU has been digitally colonized. It is a fact that the EU has been dominated by non-EU companies, especially US and Chinese firms, in the digital space. In the 2019 Forbes Top 20 digital companies list (Forbes 2019), only one EU company (Deutsche Telekom) made it to the list, while US companies claimed 12 spots; China and Japan two each; and Hong Kong, South Korea, and Taiwan one each. Likewise, in Artificial Intelligence (AI), the EU is lagging behind both the US and China, in terms of private investment and adoption of AI technologies by the private sector and by the public sector (European Commission 2018; Castro, McLaughlin & Chivot 2019). Nevertheless, this is only part of the picture.

Over the last years, the EU has responded to the growing economic and political importance of the digital economy, as well as to the security concerns of its citizens, by launching a series of regulatory initiatives (Hobbs 2020, 47). To begin with, the EU launched the Digital Single Market in 2015, to reduce barriers to digital activity



between the member-states and improve access to online services and products for citizens and businesses. After the 2013 revelations by Edward Snowden, of significant US government surveillance of European citizens' communications, including German Chancellor Angela Merkel's mobile phone, trust in the US took a significant blow and raised serious concerns regarding the cohesion of the transatlantic partnership. Snowden's global surveillance revelations triggered the debate about data protection within the EU (Rossi 2018). As a result, the EU passed the General Data Protection Regulation - GDPR. This privacy legislation imposed strict conditions on the handling of EU citizens' personal information, even if that data or citizen was physically outside the EU. When it came into effect in May 2018, companies around the world found themselves having to comply with GDPR. As a result, the EU is regarded as a standard setter in privacy and data protection, since many countries have incorporated GDPR provisions in their national legislation (Madiaga 2020, 3). Although creating EU digital sovereignty was rarely mentioned at the time, both the Digital Single Market plan and GDPR were clearly intended to enhance EU digital capabilities and provide citizens with a form of control, over their own personal data (Hobbs 2020, 47). Since then, the idea of greater European sovereignty over the digital realm has gained more ground. Indicative of the above is the reference made by Ursula von der Leyen in her statement over her policy priorities, where she called for the EU to "achieve technological sovereignty in some critical technology areas" (Von der Leyen 2019). Likewise, the European Commission stressed the importance of technological sovereignty, and the need to ensure that the EU has a secure, high-quality digital infrastructure and the ability to develop and sustain key cutting-edge technologies (Hobbs 2020, 48).

In defence of Brussel's ability to exercise its influence as a global regulatory power in the digital domain, we must stress that the EU is perceived to be a global leader in establishing standards related to online activities that are intended to safeguard its citizens and ensure an ethical approach to the dilemmas posed by the digital world (Christakis 2020, 17-20; Hobbs 2020, 49). The "right to be forgotten" and restrictions regarding hate speech are two examples of this trend. Based on the European Commission's voluntary Code of Conduct on Countering Illegal Hate Speech, many digital companies, including Facebook, Twitter, YouTube and Microsoft, adopted measures to control the speech that appears on their platforms (Bradford 2020; 132). Likewise, on February 2020, the European Commission issued three documents, a declaration about shaping Europe's digital future, a white paper on AI and the European strategy for data. These documents include rules, which ensure that data collected and controlled within the EU, is managed according to ethical standards that place privacy in the epicenter. The EU's approach towards AI is a human centric one, meaning that the EU requires compliance with fundamental rights, regardless of whether these are explicitly protected by EU treaties, such as the Treaty on European Union or by the Charter of Fundamental Rights of the European Union. Likewise, the Digital Services Act, proposes rules intended to reinforce European norms on content, consumer protection, and platform liability. In parallel, the European data strategy that was adopted in February 2020, aims to create European data spaces that will be used for economic and societal reasons (European Commission 2020). In these data spaces, EU companies and citizens will be able to control their data. By emphasizing on infrastructure, key industries, creation of data spaces and by promoting a set of norms for responsible state behavior in the digital world, the EU aspires to gain more control over how digital activities are conducted within Europe and therefore how its citizens are treated in the digital realm.

#### **4. Digital geopolitics: There is a global race for technological leadership and the EU is falling behind**

In sharp contrast to the popular belief that Europe is a rather weak player in the digital domain, the above section has illustrated the ability of the EU to exercise its power in cyberspace and regulate its digital sphere. Nevertheless, this is only one aspect of digital sovereignty. Another aspect and an equally important one, is that of digital geopolitics. The latter involves not only the politics of digital platforms that privileges certain technological giants, and the competition over the control of data, but also the division between liberal powers - US and the EU - and authoritarian ones - China and Russia - in relation to Internet freedoms and cyberspace governance. Trapped between the US Cloud Act and Chinese 5G providers, Brussels needs to balance between data localization and techno-nationalism on the one hand, and the lack of a strong industrial and technological base on the other hand (European Commission 2019).

In 2019, the EU expressed its concern about the potential reliance of its member-states on Chinese 5G infrastructure (NIS Cooperation Group 2019). Even though, Huawei was not banned, despite the pressure exercised by the US, certain member-states restrained Huawei's role in their networks (Morris 2020). The EU is concerned about the lack of control over data produced in its territory. The global cloud market is dominated by

US and Chinese technological giants. Both governments and the private sector in the EU, are concerned about using non-European data services, given the expansive extra-territorial ability granted to US law enforcement agencies to obtain foreigners' personal data under the 2018 US Cloud Act (Madiega 2020, 4). As a result, the European Commission highlighted the need to deploy European designed cloud solutions (Nextcloud 2019) and began discussions with the German and French governments, which had already launched the GAIA-X cloud project. Such initiatives aim to build a resilient digital infrastructure.

As mentioned above, technological giants like GAFAM, are collecting massive amounts of personal data and their economic model - data capitalism - is largely based on the collection and exploitation of online users' data to generate profit (Madiega 2020, 3-4). Most EU citizens store their data with US cloud providers because there are hardly any European alternatives. This is problematic and has raised concerns within the EU, because US intelligence and law enforcement agencies can access this data under the US Cloud Act. Thus, the European Court of Justice overturned in July 2020 the so-called "Privacy Shield" agreement, which allowed data transfers between European and US companies, but without providing the legal protection in the US that users enjoy in Europe. Because this data could be tapped by US authorities without EU citizens being able to take effective action against it, the Court declared it invalid.

This development is regarded, as a step towards digital sovereignty because the EU had stood up for its values and the rights of its citizens (Grüll 2020). A European alternative to the US providers is on the way. GAIA-X, a Franco-German project, is to produce cloud services according to European standards next year. It is a platform where customers can find providers that meet certain criteria, such as compliance with the GDPR. By building cloud services, the EU seeks to keep in Europe data generated on the continent and to protect that information from foreign governments (Burwell & Propp 2020, 9). US companies are also welcomed to participate, as long as they comply with these standards (Grüll 2020). This is an example of how Europe can extend its digital sovereignty, through clear sets of criteria, which companies must meet in order to be allowed to enter the internal market. Some EU member-states, including Belgium, Bulgaria, France, Germany, Greece, Luxembourg, the Netherlands, Poland, Romania, and Sweden have taken a further step, by enacting data localization measures that exclude certain categories of data from being relocated outside their territory (Burwell & Propp 2020).

Over the last three years, Europe has been discussing whether to commission Chinese producers like Huawei to equip Europe with 5G technology. The choice of 5G operators, infrastructures and their suppliers is directly linked to national security and sovereignty. Any decision about 5G cannot be made solely on terms of quality and price (Duchâtel & Godement 2019). Even though, Chinese companies offer high quality at a low price, there is a concern that the Chinese government could influence companies like Huawei to monitor or even shut down critical infrastructure whenever it wants (Grüll 2020). The US sanctioned the company and demanded Europe to follow. Brussels left the decision to the states. For example, Spain hired Huawei, whereas the Czech Republic decided not to. Germany took the middle way, welcoming all companies as long as they adhere to a catalogue of safety criteria. For example, suppliers must give a declaration of confidence that no information will reach foreign authorities and that they can refuse to disclose confidential information from or about their customers to third parties (Grüll 2020). Last, but not least, we should bear in mind that there are also credible European solutions, like Ericsson and Nokia, that the EU needs to consider (Duchâtel & Godement 2019, 18-19).

## **5. Conclusion**

In common with the early days of the Cold War, Europe is experiencing a superpower squeeze. This digital superpower squeeze places the EU between the emergent China and the US, which is struggling to retain its technological advantage. A global race for digital supremacy that includes AI and quantum computing and influences national security, global trade, and civil society, is already underway, and the EU despite its many assets, is lacking behind. It is in this environment that the debate on Europe's digital sovereignty is much needed (Christakis 2020; Hobbs 2020).

The purpose of the above analysis was not to identify a course of action for Brussels, but rather to highlight the political dilemmas that the EU is facing (European Commission 2019). The EU has launched many policies and instruments over the last years, but a wide arsenal of policy tools remains at its disposal. Building a strong industrial and technological base in the digital sector, bolstering its digital diplomacy, and improving its cyber-resilience are top priorities. Issues like the Huawei 5G offer highlight the need to develop policies that will

strengthen Europe's technological competitiveness (Duchâtel & Godement 2019). Above all, the EU needs to crystallize a strategic vision, about its future. To do that, the EU must first acknowledge its vulnerabilities and value the risks and controversies of becoming a digital fortress. Choosing between data localization and anti-trust policies and risking a technological / data war with either the US or China, is not an easy decision to make. Geopolitical antagonisms will only escalate as the digital domain becomes critical to an increasing number of actors and thus Europe has to make a meaningful and concrete decision.

## Acknowledgements

This work has been partly supported by the University of Piraeus Research Center.

## References

- Aalberts, T. (2016) "Sovereignty", in Berenskoetter, F. (ed) *Concepts in World Politics*, Sage, London.
- Barlow, J.P. (1996) "Declaration of the Independence of Cyberspace", available at <https://www.eff.org/cyberspaceindependence>.
- Betz, D.J. & Stevens, T. (2011) *Cyberspace and the State: Towards a Strategy for Cyber Power*, Routledge, Oxford.
- Biersteker, T.J. & Weber C. eds. (1996) *State Sovereignty as a Social Construct*, Cambridge University Press, Cambridge.
- Bradford, A. (2020) *The Brussels Effect. How the European Union rules the World*, Oxford University Press, New York.
- Burwell, F. & Propp, K. (2020) "The European Union and the Search for Digital Sovereignty: Building Fortress Europe or preparing for the New World?", Atlantic Council, *Future Europe Initiative, Issue Brief*, June, available at <https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/the-european-union-and-the-search-for-digital-sovereignty/>
- Carrapico, H. & Barrinha, A. (2017) "The EU as a Coherent (Cyber)Security Actor?", *Journal of Common Market Studies*, Vol. 55, No. 6, pp. 1254-1272.
- Castro, D., McLaughlin, M. & Chivot, E. (2019) "Who is winning the AI Race: China, the EU of the United States?", *Center for Data Innovation*, available at <https://euagenda.eu/publications/who-is-winning-the-ai-race-china-the-eu-or-the-united-states>
- Christakis, T. (2020) "'European Digital Sovereignty': Successfully Navigating Between the 'Brussels Effect' and Europe's Quest for Strategic Autonomy", Multidisciplinary Institute on Artificial Intelligence/Grenoble Alpes Data Institute, available at <https://ssrn.com/abstract=3748098>.
- Christou, G. (2019) "The collective securitization of cyberspace in the European Union", *West European Politics*, Vol. 42, No. 2, pp.278-301.
- Cornish, P. (2015) "Governing cyberspace through constructive ambiguity", *Survival*, Vol. 57, No. 3, pp. 153-76.
- DeNardis, L. (2014) *The Global War for Internet Governance*, Yale University Press, New Haven.
- Duchâtel, M. & Godement, F. (2019) "Europe and 5G: the Huawei Case", Policy Paper, Institut Montaigne, Paris, June, available at <https://www.institutmontaigne.org/en/publications/europe-and-5g-huawei-case-part-2>
- European Commission, (2018) "USA-China-EU plans for AI: where do we stand?", *Digital Transformation Monitor*, available at [https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/DTM\\_AI%20USA-China-EU%20plans%20for%20AI%20v5.pdf](https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/DTM_AI%20USA-China-EU%20plans%20for%20AI%20v5.pdf).
- European Commission, (2019) "Rethinking Strategic Autonomy in the Digital Age", *EPSC Strategic Note*, Issue 30.
- European Commission, (2020) "A European strategy for data", available at [https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf).
- Forbes, (2019) "Top Digital Companies – 2019 Ranking", available at <https://www.forbes.com/top-digital-companies/list/>.
- Gourley, S.K. (2013) "Cyber sovereignty", in Yannakogeorgos, P. & Lowther, A. (eds), *Conflict and cooperation in cyberspace*, Taylor & Francis, New York.
- Grüll, P. (2020) "Geopolitical Europe aims to extend its sovereignty from China", EUACTIV.DE, 11 September, available at <https://www.euractiv.com/section/digital/news/geopolitical-europe-aims-to-extend-its-digital-sovereignty-versus-china/>.
- Hobbs, C. (2020) "Europe's Digital Sovereignty: From Rulemaker to Superpower in the Age of US-China Rivalry", *European Council on Foreign Relations*, available at [https://ecfr.eu/publication/europe\\_digital\\_sovereignty\\_rulemaker\\_superpower\\_age\\_us\\_china\\_rivalry/](https://ecfr.eu/publication/europe_digital_sovereignty_rulemaker_superpower_age_us_china_rivalry/)
- Kamasa, J. (2020) "Internet Freedom in Retreat", *CSS Analyses in Security Policy*, No.273, ETH, Zurich.
- Kapsokoli, E. (2020) "EU cybersecurity governance: A work in progress", in Bellou, F. & Fiott, D. (eds) *Views on the progress of CSDP*, ESDC 1<sup>st</sup> Summer University Book, Luxembourg.
- Krasner, S.D. (1999) *Sovereignty: Organized hypocrisy*, Princeton University Press, Princeton.
- Leonard, M. & Shapiro, J. eds. (2019) "Strategic Sovereignty: How Europe can regain the capacity to act", available at [https://ecfr.eu/archive/page/-/ecfr\\_strategic\\_sovereignty.pdf](https://ecfr.eu/archive/page/-/ecfr_strategic_sovereignty.pdf)
- Liaropoulos, A. (2017) "Cyberspace governance and state sovereignty", in Bitros, G.C & Kyriazis, N.C. (eds) *Democracy and an Open-Economy World Order*, Springer, Heidelberg.
- Madiega, T. (2020) "Digital Sovereignty for Europe", *EPRS - European Parliamentary Research Service*, available at [https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS\\_BRI\(2020\)651992\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI(2020)651992_EN.pdf)

**Andrew Liaropoulos**

- Morris, I. (2020) "Europe is showing Huawei the exit", *Light Reading*, 9 September, available at <https://www.lightreading.com/5g/europe-is-showing-huawei-exit/d/d-id/763814>
- Mueller, M.L. (2010) *Networks and States: The Global Politics of Internet Governance*, MIT Press, Cambridge.
- Mueller, M.L. (2020) "Against Sovereignty in Cyberspace", *International Studies Review*, Vol. 22, No. 4, pp.779-801.
- Nextcloud, (2019) "EU governments chose independence from US cloud providers with Nextcloud", 27 August, available at <https://nextcloud.com/blog/eu-governments-choose-independence-from-us-cloud-providers-with-nextcloud/>
- NIS Cooperation Group, (2019) "EU coordinated risk assessment of the cybersecurity of the 5G networks", Report, 9 October, available at [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_19\\_6049](https://ec.europa.eu/commission/presscorner/detail/en/IP_19_6049)
- Rossi, A. (2018) "How the Snowden Revelations Saved the EU General Data Protection Regulation", *The International Spectator*, Vol. 53, No. 4, pp. 95-111.
- Slomp, G. (2008) "On Sovereignty", in Salmon, T.C. & Imber, M.F. (eds) *Issues in International Relations*, Routledge, New York.
- Timmers, P. (2019) "Strategic Autonomy and Cybersecurity", EU Cyber Direct, Supporting EU Cyber Diplomacy, available at [https://eucyberdirect.eu/content\\_research/strategic-autonomy-and-cybersecurity/](https://eucyberdirect.eu/content_research/strategic-autonomy-and-cybersecurity/)
- United Nations General Assembly, (2015) "Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security", available at <https://dig.watch/sites/default/files/UN%20GGE%20Report%202015%20%28A-70-174%29.pdf>
- Von der Leyen, U. (2019) "A Union that strives for more", Political Guidelines for the next European Commission 2019 - 2024, October 9, available at <https://op.europa.eu/en/publication-detail/-/publication/43a17056-ebf1-11e9-9c4e-01aa75ed71a1>
- Wu, T.S. (1997) "Cyberspace Sovereignty? - The Internet and the International System", *Harvard Journal of Law and Technology*, Vol. 10, No. 3, pp. 647-666.

# Mandatory Cybersecurity Training for all Space Force Guardians

**Banks Lin, Mark Reith and Wayne Henry**

**Air Force Institute of Technology, Wright-Patterson Air Force Base, USA**

[banks.lin@afit.edu](mailto:banks.lin@afit.edu)

[mark.reith.ctr@afit.edu](mailto:mark.reith.ctr@afit.edu)

[wayne.henry@afit.edu](mailto:wayne.henry@afit.edu)

DOI: 10.34190/EWS.21.107

**Abstract:** The most vulnerable and exploitable aspect of a computer network is often the user. Each user that operates an endpoint machine or system on the network increases the enterprise footprint for adversaries to target. Users introduce human error caused by the lack of knowledge or poor training on a particular tool or application. To date, the United States Air Force has not provided fundamental cybersecurity training for all its personnel. Currently, Airmen can go through basic or officer training and enter active duty without compelling training on cybersecurity. It is not until after airmen are sworn in that they are mandated to complete a computer-based cyber awareness training. This exposure to cybersecurity is essential but insufficient. This article proposes an enhanced level of mandatory cybersecurity education and training during basic and officer training to normalize conversations on cybersecurity to create an informed and well-educated United States Space Force. This ensures all Guardians understand the cyber vulnerabilities and threats to the point of it being common knowledge. Further, we use evidence-based cybersecurity training techniques to develop a course plan and learning objectives for Guardians to better retain and apply cybersecurity knowledge. This plan integrates realistic practice, cyber expertise, and personal reflections.

**Keywords:** mandatory cybersecurity training, evidence-based, annual refresher, space force

---

## 1. Introduction

The Government Accountability Office (GAO) released a report in 2020 stating that the Department of Defense (DoD) needs to improve on cyber hygiene. The report outlines how the cyber programs to train and equip do not match the severity of the problem (Government Accountability Office, 2020). Still, the military's most secretive weapon systems that impose the strictest security measures require humans to operate and maintain. It is widely known that users are the biggest threat to any computer system or network (Beyer and Brummel, 2015). The average user in an organization does not possess the ability to differentiate a malicious event or a computer error encountered on the daily basis. For this reason, it is imperative that all users must learn cyber vulnerabilities and threats to deter them from doing something that may compromise their machine and the entire system network.

Cybersecurity training must be required for all users. Organizations should impose a culture where cybersecurity is second nature or common knowledge, much like humans locking their doors when not actively using them when living in a bad neighborhood. To propose mandatory cybersecurity training for the Space Force, we will discuss what it will look like, who it would apply to, when and where it will be held, why it is better, and how it will be implemented. More importantly, we will focus on implementing evidence-based training to ensure the delivery of the training that can be better retained and be put into practice.

For this paper, the scope will be limited to the US Space Force because it has the potential to become the pathfinder and establish a cybersecurity-focused culture. In the Space Force, its users are called Guardians. The Chief of Space Operations, General Raymond, has set the vision for the Space Force to be lean and agile, shifting the culture that puts "speed, technology, and innovation" at the forefront of this "digital service" (Pope, 2020). Based on General Raymond's vision, it is likely that the newest service will heavily rely on the cyberspace domain to operate its mission. Therefore, it is even more critical to instill mandatory cybersecurity training. The Space Force is the perfect candidate for this training; though, this can be applied to the rest of the DoD and the federal government. DoD must include and integrate cybersecurity awareness plans in future continental US and overseas military operations.

## 2. Characterizing the current challenge/problems

Cyber professionals and hackers would agree that users are the most vulnerable point of a computer network. The larger the organization, the more users there are that have accounts with access to permissions to the system. The more users there are, the larger the footprint of the organization which increases the number of

attack vectors and attack paths an adversary can leverage. Without proper training for all network users, the adversary would have the advantage since they just need to find the weakest link to exploit and gain a foothold on the network. Attackers can leverage naïve users as an access point to escalate into the network and potentially monitor, steal, or exfiltrate sensitive data. Currently in the Air Force, and the Space Force, cyber awareness and computer-based training are two of the cyber-related topics out of the total 161 topics covered in Basic Military Training (Air Force, 2020). This leaves a total amount of cybersecurity exposure to about 1.2% of the eight and a half week training which is not nearly enough for these young Guardians. Cybersecurity is only taught to certain career fields or “job-specific” assignments that are exposed to the cyberspace domain. In *Improving the Pipeline*, the authors focused on bringing cyber education to high school education to create more cyber professionals into the workforce due to the shortage of these experts internationally and in the United States. However, they concluded that even though the course curriculum for these high school students were successful, the better course of action was to increase cybersecurity awareness and training for *more* individuals, a mile-wide and inch deep concept (Gorka et al., 2020). In other words, it may be important to train and equip Airmen and Guardians to tackle the cybersecurity challenges we face, but the DoD should not solely rely on these cyber warriors to solve these challenging problems. It takes the whole force to be trained and equipped to exercise good cyber hygiene and best practices to uphold the security posture and defense of the system. Cybersecurity is not a challenge that *they* can solve. It is a problem that the *total force* must solve. The complexity of the program is quite daunting and a challenge to manage and overcome. However, normalizing the understanding of cybersecurity will equip people with the necessary tools to tackle the problem head on. It starts at the grassroots - better known as the everyday users.

The space domain plays a critical role in our nation’s defense. Military use of space assets includes anti-ballistic missiles, communications, navigation, and missile warning. Commercial use of space assets is seen in our iPhones, Google Maps, or Uber. These systems utilize global positioning systems (GPS) to detect and determine location, navigation, tracking, mapping, and timing. What happens when we can no longer trust the information provided by the GPS? What happens when the data is no longer available or compromised? How does the military operate without its space assets? Space and cyber are two intertwined domains that depend on each other (Al-Rodhan, 2020). We depend on the integrity of the data that is pushed up to space and pulled down to our ground systems where we process the data. During this process, the information flow could be manipulated or jammed, leaving most present-day space threats to be cyber-related (Gould, 2018). The US government recognized the importance and our dependence in the space domain with the creation of a separate military service to focus on national security in space in 2018 (Loverro, 2018). In a service created to defend space against all threats, it is crucial to have personnel trained and equipped with cyber knowledge.

### **3. Current cyber education/training and its shortfalls**

Currently, cybersecurity training in the Space Force is identical to the Air Force. Guardians in the cyber career field attend an initial training pipeline to be equipped for their occupation specialty. However, there is no requirement for functional personnel, outside of the cyber career fields, to have cyber education or training prior to or after entering the workforce as active duty, reservist, or guard. For example, the Air Force does not require finance personnel, responsible for managing base finances, to have cybersecurity training other than the cyber awareness challenge. However, finance personnel have access to sensitive operational and personal information. Most of the time, finance personnel operate on the Air Force Network to perform their day-to-day operations. Their footprint on the enterprise network has the potential to be a vulnerable attack vector for an adversary. Each one of these users are humans that an attacker can exploit through weak passwords, email phishing, or social engineering.

The only cybersecurity training for all personnel required by DoD policy is the DoD Cyber Awareness Challenge, a computer-based training module used to provide user training and awareness on cybersecurity. The training focuses on cybersecurity issues caused by human error such as social engineering attacks, social media usage, and general cyber hygiene best practices. However, it falls short in its delivery because DoD personnel receive the same content and module every year with minimal changes to the course (Reith et al., 2018, pp.439-447). It is difficult to retain knowledge that is presented in unsupervised training slides that anyone can “click-through” to satisfy a requirement. We are not proposing to eliminate the Cyber Awareness Challenge. Instead, we recommend changing the annual training requirement to update the total force on the latest cyber threats and cyber breaches to the US government in the recent year. All users who partake in this training should have a foundational understanding of cybersecurity prior to taking this computer-based training. The Cyber Awareness

Challenge should only be used as a refresher to ensure all personnel are up to date on the cutting-edge advances in the cyberspace domain. It should be treated like a booster vaccine shot that re-exposes and increases immunity against a particular antigen, the same way the Cyber Awareness Challenge should re-expose students to a specific topic in detail. However, this computer-based training is being used as the prime dose of a vaccine shot to provide the broad, encompassing dose of education that covers a redundant number of topics without depth.

The bulk of Air Force cybersecurity training opportunities are for personnel in the cyber career field. Cyber airmen go through an initial training pipeline at Keesler Air Force Base to learn the fundamentals of cybersecurity and graduate with a commercial IT certification, Security+ (Reith et al., 2018, pp.439-447). The Air Force pushes out over 5,000 cyber airmen annually. Many of them follow-on this training with initial qualification training or mission qualification training to receive weapon system specific training. The Air Force even offers graduate-level education on many STEM degrees, to include Cyber Operations, at the Air Force Institute of Technology. Air Force and Space Force officers also have unique opportunities to attend prestigious and highly rigorous training programs such as Cyber Weapons School and the National Security Agency sponsored Computer Network Operations Development Program. These programs train airmen on cutting-edge cyber technologies, techniques, and procedures to lead, protect, and defend this nation in the cyberspace domain. There is no doubt that the Air Force and Space Force possess the ability to educate its Airmen and Guardians into cybersecurity experts. However, these opportunities are largely for the cyber career field airmen excluding the rest of the force. This leaves room for improvement for every other career field to learn and understand the fundamentals of cybersecurity.

The Space Force has an opportunity to drastically change the culture and dynamic in how the service views the cyberspace domain. Understanding the fundamentals of cybersecurity can become a norm, a standard, or even a requirement. Guardians must understand the fundamental concepts in cybersecurity the same way they understand the importance of dress and appearance. The cyberspace domain is an intrinsic part of military operations, our society, and our daily lives. Therefore, it is critical that Guardians understand the potential threats and vulnerabilities in this domain that the nation depends on.

#### **4. Proposed strategy and framework**

Users expand the footprint of the network which allows attackers to leverage them as potential attack vector for exploitation (Anderson, 2020). What are some ways to limit these attack vectors on the network while allowing users to operate and continue their mission? The proposed solution is twofold: to have the Space Force train all Guardians on the fundamentals of cybersecurity prior to entering the service, and to include annual refresher or “booster” courses to provide detailed exposure on a specific topic. To execute these two proposed approaches, we must implement an evidence-based framing methodology.

##### **4.1 Mandatory cybersecurity training**

We propose mandatory cybersecurity training to be implemented for all Guardians during basic and officer training, prior to them entering the workforce. With our increasing reliance on technology, the US cannot ignore the growing cyber threats. Instead, action must be taken to limit potential impacts to network security by properly training and equipping the human user (Corren, 2020). Our proposal highlights cybersecurity training during basic and officer training because imposing cyber awareness can be difficult when it is conducted after Guardians are already in the workforce. Cyber development programs to improve awareness to the functional community (non-cyber) in the Air Force are too technical or generic to be memorable and engaging. Commanders also face resource and logistical barriers to improve on their troop’s cyber knowledge (Reith et al., 2018, pp.439-447). Basic training and Officer Training School can be readjusted to incorporate an hour out of the training day to take a cybersecurity course. The US Air Force Academy and Reserve Officer Training Corp can add a course requirement for one quarter or semester to take a cybersecurity course to complete before graduation. Currently, military personnel go through basic or officer training to learn military bearing, basic land navigation, and physical fitness as part of the training process to transition from civilian to military life. With an emphasis on a “digitized” force, the Space Force should focus its basic training on mental discipline, the same way the Army focuses on the physical discipline. For Guardians, cybersecurity should be part of the curriculum and acknowledged as a minimum standard to be part of the Space Force. Cybersecurity should be common knowledge the same way members in the military considers Dress & Appearance. Implementing improved

cybersecurity training early in a career is more relevant than ever and this paradigm shift in culture must start in the newest service branch.

## 4.2 Evidence-based framing strategies

Since every trainee or cadet does not have a technical background, a solution is to employ evidence-based framing training. Evidence-based framing focuses on the message framing of sociotechnical interaction between people and cybersecurity in the workplace. This would ensure Guardians retain the knowledge and apply it to their work to diminish the vulnerabilities exposed by human users. The authors of *Building Cybersecurity Awareness: The need for evidence-based framing strategies* states that evidence-based framing has two main goals: 1) frames are based on facts and 2) facts must be presented in good messaging frames. By implementing this training strategy, the network users will benefit with the understanding and awareness of the complex challenges in cybersecurity. Message framing is a communication strategy to turn complex problems into clear and easily explainable problems. Message framing must be able to carry complex reality into a simple framing and presented in a way that cannot be rebuked. Table 1 summarizes six effective strategies to convey the importance of cybersecurity (de Bruijn and Janssen, 2017).

**Table 1:** Summary of evidence-based framing strategies (de Bruijn and Janssen, 2017)

<i>Strategy</i>	<i>Description of an effective frame</i>
1. Do not aggravate cybersecurity	Do not give a doom and gloom pitch on cybersecurity. Provide realistic perspective.
2. Clearly identify the adversaries	Identify who the adversaries are and clearly identify what is considered legal and illegal
3. Highlight blue suiters and US cyber capability	Highlight the success of US cyber capabilities and what they have accomplished
4. Demonstrate importance for the Space Force mission and ultimately the national security	Show impact of what due diligence from every individual can provide from a strategic view (i.e., economic growth and prosperity from cyberspace dominance)
5. Build and analyze scenarios and identify corrective action in case studies	Build correct/incorrect scenarios to bring practical experience; identify improper cases to learn how to perform better in the future
6. Connect cybersecurity to aspects of their personal lives	Provide understanding that cybersecurity is an integral part of everyone’s lives but may not receive enough attention

### 4.2.1 Do not aggravate cybersecurity

The primary goal of not aggravating cybersecurity is to prevent denial or the lack of enthusiasm. For example, lectures that cyber threats are everywhere and that everything done on the web is being watched can lead students to believe an adversary can handily defeat our systems. Although these threats exist, they are certainly exaggerated to the point where the students may react in denial. This frame can lead students to believe the threats “won’t happen to me” or “this is just a class we have to get through”. For an effective frame, the course curriculum must have a realistic perspective of the topic. During Space Force basic and officer training, do not leave case studies or examples of cyber threats open-ended without a solution. It is important to bring up case studies to demonstrate real-world examples of cyber threats but include a solution for students to understand what the countermeasures are to combat it in the future.

### 4.2.2 Clearly identify the adversaries

The cyberspace domain is unique to the other domains because it is difficult to identify exactly who the enemy is or attribute an attack to a specific nation-state, organization, or group. Without a clearly defined adversary, it is difficult to frame cybersecurity in an effective way for the audience. In cyberspace, the attribution of various attacks is usually in the form of an IP address or based on the level of sophistication of the attack. From the Space Force perspective, describing the real threats posed by certain nation-states, rogue actors, or state-influenced actors may help the students better frame the cybersecurity landscape. The course curriculum must also explain the differences between the cyberspace domain compared to the other domains and why that changes the dynamic of cyber warfare.



**4.2.3 Highlight blue suiters and US cyber capability**

Highlighting the “good guys” during cybersecurity training is just as important as identifying the “bad guys”. The US government is widely understood to have the most sophisticated cyber-weapons and is expanding its cyber capability through offensive cyber operations to its military organizations (Whyte and Mazanec, 2019, pp.261–262). Most Americans are unaware of who is protecting US interests in cyberspace, including fresh recruits joining the Space Force. By highlighting the strategic vision of Space Delta 6, the Space Force component responsible for cyberspace operations, and the cyberspace defenders, we clearly frame our national defensive capabilities in the environment. The next time the server goes down at the office, the students may have better appreciation for the communication support team working tirelessly to get servers back up and running.

**4.2.4 Demonstrate importance for the Space Force mission and ultimately national security**

Leadership makes a decision that is in the best interest of the organization when presented with all the possible information. However, Guardians at the bottom of the chain of command may not see the strategic vision behind that decision and may become unresponsive to the vision. This framing strategy emphasizes the importance of tying the course content being taught to the overarching strategic vision of the Space Force mission and the long-term effects on national security for the US.

**4.2.5 Build and analyze scenarios and identify corrective action in case studies**

This framing strategy focuses on building scenarios that can capture the attention of the students while being relevant in a real-world context. The scenarios should include appropriate and non-appropriate actions to enhance the learning experience for the students. Case studies allow students to study and analyze the event that have taken place in the past or that are still ongoing. When students identify corrective actions in case studies, it encourages them to use their critical thinking skills to solve the problem and potentially inform others to understand how to perform or change things for the better in the future. Space Force Guardians in basic and officer training can learn about case studies such as the events that occurred with Stuxnet or Edward Snowden.

**4.2.6 Connect cybersecurity to aspects of their personal lives**

The last framing strategy is capturing the attention of the audience. It is difficult to create content that is relatable to someone who does not have a technical background. The training course must be engaging to ensure students can retain the knowledge and apply it when they graduate and enter the workforce. The Space Force could implement a capstone project as an end-of-course exercise during basic and officer training to instruct students how to crack a password. A simple step-by-step instruction to crack a password with real hacking tools can give a tangible feeling to hacking a password. The intent would not be to develop elite hackers, but to expose how easy it can be to hack an account.

Table 2 provides the proposed cybersecurity training course objectives, including what should be the expected knowledge for all Guardians. By following these evidence-based framing strategies, Guardians will not be intimidated by the potentially complicated and challenging cybersecurity threats. Correctly framing the message will improve the communication with students who do not have the technical background. Training should include hands-on experience in a relevant cybersecurity area of sufficient complexity. Understanding how easy it can be to crack a password and compromise a machine on a network can improve the students’ cybersecurity posture throughout their career. Investing in mandatory cybersecurity training is investing in a well-equipped and organized Space Force.

**Table 2:** Course objectives

<i>Topic</i>	<i>Content</i>
History of cybersecurity	Provides why information dominance is important to Space Force and national security
Case studies	Provides how cyber can impact space mission in real-world scenarios
High-level cyber terms	Exposes high-level cyber terms (i.e., IP, encryption, digital signature) common in space systems
Most common vulnerabilities	Identify and prevent common cyber vulnerabilities (i.e., jamming, phishing, social engineering, weak passwords) in space systems
Capstone	Conduct an exercise for students to go through the process of cracking a password

### 4.3 Annual refresher training

Once Guardians enter the workforce with their foundational cybersecurity training, it is important to review this knowledge on an annual basis since cyberspace is an everchanging domain. A prime-boost immunization strategy is one of many different vaccine modalities that could enhance the immunity in someone. The general concept of this strategy has two types of doses: prime and booster. The prime dose would inject into the immune system with an immunogen and the booster dose injects with a different immunogen. The purpose of the prime-boost approach is to develop greater immunity compared to a single dose vaccination or multiple doses with the same antigen (Valdés et al, 2019). In the prime-boost immunization strategy example, the foundational cybersecurity training would be considered the prime dose vaccine shot while the annual refresher would be considered the booster shot. The annual refresher should cover unclassified intelligence reports on the latest cyber threats or case studies on recent cyber breaches in the US government to generate awareness. The annual refreshers can also conduct random cyber awareness testing, like the current Cyber Awareness Challenge, but cover a limited number of topics in greater detail such as a year designated to phishing and another year for weak passwords. In the Space Force, only cyber professionals receive both the prime and booster shots of cybersecurity training. Instead, prime and booster should be applied to all Guardians to ensure members in the Space Force are fully inoculated against poor cyber hygiene practices. With the proposed approaches of mandatory cybersecurity training and an annual refresher, we believe the Space Force will develop good cyber hygiene to combat the cyber threats. It is critical that the US maintains information dominance. The Space Force and its assets can only be as strong in cybersecurity as its weakest link.

## 5. Conclusion

Today's modern warfare is no longer like depictions found in video games like "Call of Duty". Instead, modern warfare is conducted in the cyberspace domain. This domain is especially relevant for the Space Force because its mission depends on satellite communications, information technology, and a reliable enterprise network. Cybersecurity experts can help identify the technical vulnerabilities that the United States faces in defense. However, few can point to a solution that is realistic and obtainable in this evolving sociotechnical landscape. Evidence-based cybersecurity training during basic and officer training could fundamentally change the culture of the services within the DoD and eventually the federal government. It can begin to normalize conversations of cybersecurity, rather than treating it as something that the cyber experts will figure out. A cultural change is necessary to realize the necessity and importance of understanding cybersecurity for all Guardians, regardless of military branch or job function. The culture change will need to start from the grassroots and start small. What better way to implement this new change to the nation's newest military branch, the US Space Force?

**Disclaimer:** The views expressed are those of the author and do not reflect the official policy or position of the US Air Force, US Space Force, Department of Defense, or the US Government.

## References

- Air Force. (2020) *Basic Military Training* [online]. Available at: [https://www.airforce.com/pdf/USC91019023\\_BMT\\_Schedule.pdf?\\_ga=1864951899.1609641397](https://www.airforce.com/pdf/USC91019023_BMT_Schedule.pdf?_ga=1864951899.1609641397) (Accessed: 31 December 2020)
- Al-Rodhan, N. (2020) *The Space Review: Cyber security and space security* [online]. Available at: <https://www.thespacereview.com/article/3950/1> (Accessed: 10 December 2020)
- Anderson, F. (2020) *5 Reasons Why Your Employees are Your Biggest Cybersecurity Threat* [online]. Available at: <https://blog.symquest.com/why-human-error-biggest-cyber-security-vulnerability> (Accessed: 8 November 2020)
- Beyer, R. and Brummel, B. (2015) *Implementing Effective Cyber Security Training for End Users of Computer Networks* [online]. Available at: <https://www.shrm.org/hr-today/trends-and-forecasting/special-reports-and-expert-views/Documents/SHRM-SIOP%20Role%20of%20Human%20Resources%20in%20Cyber%20Security.pdf> (Accessed: 8 November 2020)
- Corren, A. (2020) *Cyber Security Basics All Employees Must Be Trained in* [online]. Available at: <https://trainingmag.com/cyber-security-basics-all-employees-must-be-trained/> (Accessed 11 December 2020)
- de Bruijn, H. and Janssen, M. (2017) *Building Cybersecurity Awareness: The need for evidence-based framing strategies* [online]. Available at: <https://www.sciencedirect.com/science/article/pii/S0740624X17300540> (Accessed: 10 December 2020)
- Gorka, S., McNett, A., Miller, J. and Webb, B. (2020) 'Improving the Pipeline', *Journal of The Colloquium for Information Systems Security Education*, 7(1), p. 5.
- Gould, J. (2018) *Think Space Force is a joke? Here are four major space threats to take seriously* [online]. <https://www.defensenews.com/space/2018/08/09/think-space-force-is-a-joke-here-are-four-major-space-threats-to-take-seriously/> (Accessed: 8 November 2020)

**Banks Lin, Mark Reith and Wayne Henry**

- Government Accountability Office. (2020) *CYBERSECURITY DOD Needs to Take Decisive Actions to Improve Cyber Hygiene Report to Congressional Committees United States Government Accountability Office* [online]. <https://www.gao.gov/assets/710/705886.pdf> (Accessed: 10 December 2020)
- Loverro, D. (2018) *Why the United States needs a Space Force* [online]. Available at: <https://spacenews.com/why-the-united-states-needs-a-space-force/> (Accessed: 10 December 2020)
- Pope, C. (2020) *Driven by 'a tectonic shift in warfare' Raymond describes Space Force's achievements and future* [online]. Available at: <https://www.spaceforce.mil/News/Article/2348423/driven-by-a-tectonic-shift-in-warfare-raymond-describes-space-forces-achievements/> (Accessed: 11 December 2020)
- Reith, M., Trias, E., Dacus, C., Martin, S. and Tomcho, L. (2018) 'Rethinking USAF cyber education and training'. In *Proceedings of the 13th International Conference on Cyber Warfare and Security, ICCWS 2018*, pp. 439-447.
- Valdés, I., Lazo, L., Hermida, L., Guillén, G. and Gil, L. (2019) *Can Complementary Prime-Boost Immunization Strategies Be An Alternative And Promising Vaccine Approach Against Dengue Virus?* [online]. Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.01956/full> (Accessed: 30 December 2020)
- Whyte, C. and Mazanec, B.M. (2019) *Understanding Cyber-Warfare: Politics, Policy and Strategy*, London, New York Routledge, pp.261–262 (Accessed: 11 December 2020)

# The Challenges to Cybersecurity Education in Developing Countries: A Case Study of Kosovo

Arianit Maraj<sup>1</sup>, Cynthia Sutherland<sup>2</sup> and William Butler<sup>2</sup>

<sup>1</sup>AAB College, Faculty of Computer Sciences, Kosovo

<sup>2</sup>Capitol Technology University, Laurel, Maryland, USA

[arianit.maraj@universitetiaab.com](mailto:arianit.maraj@universitetiaab.com)

[cevalentine@captechu.edu](mailto:cevalentine@captechu.edu)

[whbutler@captechu.edu](mailto:whbutler@captechu.edu)

DOI: 10.34190/EWS.21.003

**Abstract:** Preventing cyberattacks depends on educating and training staff to acquire sufficient knowledge and skills to protect against such attacks. So far, in developing countries, there is little research focused on identifying factors that hinder the development of cybersecurity education in those countries. Therefore, studying these factors is very important. Cybersecurity is a relatively new profession and as such, suffers from a lack of standard mechanisms to bring the results of research into the curriculum and include students in academic research. The challenges to cybersecurity in developing Countries faced by Higher Educations Institutions (HEI) are huge. In general, it is not yet understood that HEIs are the core of the solution to the problems and challenges faced by Kosovo. Combining the core values of security, privacy, and HEI experimentation poses significant benefits to online security. The first step in meeting these challenges is to develop and execute an applicable education strategy. In Kosovo, there exists a cybersecurity strategy developed by the Government, which is deficient and not fully implemented or practiced. This strategy emphasizes the need to focus on providing a roadmap to develop a cybersecurity curriculum and advanced learning modules. However, so far, there is only one accredited cybersecurity academic program in Kosovo. This situation illustrates that cybersecurity education and training standards have not been given proper attention. Therefore, developing a standardized curriculum for cybersecurity is an urgent need. It is also recommended that a robust cybersecurity strategy, which focuses on cybersecurity education and training standards be developed. In this paper, the key factors in cybersecurity education and the main strategies and recommendations for advancing cybersecurity education in Kosovo will be explored and proposed.

**Keywords:** cybersecurity, education, training, awareness-raising, cyber strategy

---

## 1. Introduction

Internet security incidents are increasing at an alarming rate and may endanger essential critical infrastructure services such as water, healthcare, electricity, and other basic services. Threats can have different origins: criminal attacks for financial gain, politically motivated, terrorist, or state-sponsored. Economies around the World are already affected by cybercrime. Online criminals are using sophisticated methods to interfere in information systems, stealing critical data or holding companies hostage with ransomware. The rise of economic espionage and activities sponsored by various countries on the Internet presents a new category of threats to governments and companies around the World. Staying protected from cybersecurity threats requires that all users, from children and their parents, become more sophisticated users, to be more aware and educated continuously. However, exposure to cyber awareness at all levels of education, including elementary education, higher education, universities, and lifelong education is essential. This effort would promote education and encourage the adoption of cybersecurity competencies throughout the Country.

The main purpose of cybersecurity education is to enlighten the users of technology about the potential risks while using the Internet. When users are educated about the risks that they may encounter while communicating online, these risks can be addressed and significantly reduced (Dlamini et al). Children are prone to cyberattacks because they are curious about exploring the Internet World, but are unaware of the risks associated with using Internet communication tools (von Solms et al).

Typically, cybersecurity executives in higher education spend a small percentage of their time developing strategies, however, these activities are likely to have the greatest impact on their Institutions. Having a strategy that evolves and adapts to a changing environment can turn a good security team into a great one (EDUCAUSE). Usually, experts have tried to adapt the security strategy to an IT strategy or a business strategy approach, but it should be clear that security strategy differs greatly from the IT and business strategy. The security strategy should support and enable the IT and business strategies.

Since Education and Training for cybersecurity is a dynamic process due to its continuous evolution, we need to have a clear methodology for gaining quick insights from the currently available data sources.

The coverage is uneven across Europe, however, the availability of online security courses and training is increasing, especially in the area of core security curricula (ENISA). In developing countries, such as Kosovo, there are very few curricula related to cybersecurity education at all levels of education. This trend is also observed elsewhere in the Region.

Cybersecurity curricula should implement mechanisms to incorporate changes in technology. Currently, the lack of such mechanisms has resulted in very few materials on present and emerging threats. As a result, education provided under various cybersecurity programs tend to be organized around a useful common goal, but struggles to match the requirements of the dynamic workplace (ENISA). Some EU countries are making efforts to bring cybersecurity students into contact with industry and Government professionals to participate in practical projects. These countries have realized that developing cybersecurity curricula requires close cooperation with Industry and other Governmental Institutions.

The methodology used in this manuscript is presented in the next section. In section 3, we describe the background, including the cyber education objectives for some of the UE Countries. In section 4, we present cybersecurity education, training, and awareness in Kosovo. The key factors in cybersecurity education in Kosovo are presented in section 5. Conclusions and recommendations are drawn in the last section.

## **2. Methodology**

This manuscript consists of a comparative analysis of US, EU, and Western Balkans initiatives. These comparative analyses were conducted to provide awareness-raising, training, and educational activities designed to expose cybersecurity risks in those respective countries. The analysis of the literature regarding the effectiveness of Cybersecurity education and awareness-raising issues was also examined. This analysis has involved a detailed literature search regarding the cybersecurity strategy in some of the EU countries, where the focus was on some of the main objectives that they have implemented in a cybersecurity strategy and action plans (See Table 1). Besides this, to analyze the challenges of cybersecurity education in developing countries, we have analyzed the latest literature including here the professional reports on cybersecurity maturity level in all of the Balkans Countries; Albania, Kosovo, North Macedonia, Serbia, and Montenegro (See Table 2). For the Western Balkans, the focus was on some of the main objectives regarding cyber education. These objectives include cybersecurity education in primary and secondary schools, the higher education level, cybersecurity training, cybersecurity skills, and awareness-raising on cybersecurity. We also analyzed to identify successful cybersecurity strategies that have been implemented by other countries in the region that might be suitable for Kosovo (See Table 2).

## **3. Background**

Internet security challenges include not only technology, processes but the users themselves. HEIs should establish advanced programs to educate users to protect sensitive data and networks. The focus should be on efforts to reduce risks when using technology. HEIs are pursuing different approaches to security education. Some believe in early specialization to focus more on implementing Internet security, making it a part of general university education. Others are not advocates of specialized undergraduate degrees and feel that it is more important to establish a solid foundation inside computer science. Existing cybersecurity education programs within the region do have some limitations. However, to establish and maintain a technical advantage, cybersecurity education must deliver the latest technology and techniques taught by experts to the cyber-capable workforce. Cybersecurity education is being prioritized around the World. According to (ENISA NCSS) one of the first countries to recognize cybersecurity as a national strategic matter is the United States. In 2003, the US published the National Strategy to Secure Cyberspace (Bush, G.W). This strategy was a part of the overall National Strategy for Homeland Security (DHS) and was developed as a response to terrorist attacks on September 11<sup>th</sup>, 2001. After those events, some action plans and strategies began to emerge throughout Europe: in Germany (2005), in Sweden (2006), Estonia (2007). Since then, considerable work has been completed and almost all of the EU Countries have published a national cybersecurity strategy. By studying the strategies of some of the EU countries, the common objective of all countries is "Strengthen training and educational programs" (See Table 1). Some of the Countries, besides education, have given special importance to research and development as well (See Table 1). Based on the analysis, we can conclude that there are clear cybersecurity

strategies at the national level in most EU countries. Since Kosovo is part of the Balkan Countries, we have also studied policies related to cybersecurity strategies in the Western Balkans.

Concrete cybersecurity activities in the Western Balkans were mentioned in the 2014 activities, where the "Winter School of Youth Network Security for the Western Balkans and Moldova" was organized (KCSS). Following this event, several other activities were organized in the region, which had a direct impact on improving the cybersecurity knowledge of Western Balkans countries. Seeing the importance of Internet security, such activities have been intensified in the years that followed, always regarding cyber strategies in the US and the EU (Bush, G.W).

**Table 1:** Cyber objectives for some of the UE Countries (ENISA reports for: Italy, Croatia, Slovenia, Hungary, Finland, Slovakia, France, Austria)

Objectives/ Countries	Austria	Italy	Finnish	Slovenia	Hungary	Slovakia	France
Address cybercrime	√	√	√	√			√
Balance security with privacy	√	√	√	√			√
Citizen's awareness		√	√	√		√	√
Critical Information Infrastructure Protection	√	√	√	√	√	√	√
Engage in international cooperation	√	√	√	√	√	√	√
Establish a public-private partnership	√	√	√			√	
Establish an incident response capability	√	√	√	√	√	√	√
Strengthen training and educational programs	√	√	√	√	√	√	√
Foster R&D	√		√			√	√
Develop national cyber contingency plans		√	√				√
Organize cybersecurity exercises	√	√		√	√		√
Establish an institutionalized form of cooperation between public agencies	√	√		√	√	√	√
Establish baseline security requirements		√			√	√	√
Provide incentives for the private sector to invest in security measures		√					

Until 2016, there has been little research available to specifically address developments in the Western Balkans in the field of cybersecurity. However, some related global reports also reflect the situation in some or most of the region's countries, such as the ITU's "Global Cybersecurity Index 2015 and Cyberwellness Profile" and BSA's "Cybersecurity Maturity Panel BE 2015 " (ITU, EU Cybersecurity).

Recently, several scientific studies are stating the importance of cyber education in Western Balkans (Maraj, A, Rizmal, I, Šendelj, R et al, Poposka, V). A study of the existing security cooperation mechanisms in the Western Balkans Region is presented in (Minović, A et al). The main idea of this study was to identify gaps in and potential for cybersecurity cooperation across the region. A review of major projects and funding opportunities in cybersecurity in the Western Balkans by major international organizations is presented afterward. The cybersecurity situation in the Western Balkans is the same for all Countries.

### **3.1 Education, training and awareness in Bosnia and Hercegovina**

In October 2018 the main Institutions of Bosnia and Hercegovina, conducted a review of the nation's cybersecurity capacity (GFBH). The main objective of this review was to enable the Government to understand the national cybersecurity capacity to develop the Country's national cybersecurity strategy and to strategically prioritize investment in cybersecurity capacities. The participants in this review defined that five dimensions of cybersecurity capacity should be considered as a priority for Bosnia and Hercegovina: Cybersecurity Policy and Strategy, Cyber Culture and Society, Cybersecurity Education, Training and Skills, Legal and Regulatory Frameworks, Standards, Organizations, and Technologies. Awareness of cybersecurity risks and threats in Bosnia and Hercegovina is still low at all levels of society and awareness-raising is not a priority for government institutions. This situation persists in part due to governmental lack of knowledge of possible risks and existing threats.

### 3.2 Education, training and awareness in Serbia

The cybersecurity strategy in Serbia was adopted in 2017. The Strategy clearly defines priority areas that include the security of information and communication systems, security of citizens when using technology, fight against high-tech crime, and information security of the Country. The Strategy stated that the education system should enable knowledge acquisition in the field of information security; therefore, this strategy is a significant step forward in efforts aimed at building capacity in information security of society as a whole - from elementary school to study programs at universities (GRS, Rizmal, I). Also, the growing trend of cyber training programs within is noticed within the large companies in Serbia.

### 3.3 Education, training and awareness in Montenegro

The state of cybersecurity education in Montenegro is stated within the Roadmap for new Cybersecurity Education in Montenegro created in 2015 (Šendelj, R et al). There is only one University that offers two post-graduate study programs regarding cybersecurity. These programs in the field of cybersecurity that are being developed in this University are important forms of formal education. These programs provide a high-quality and systematic education system, needed to meet national objectives in this area. The level of awareness about cybersecurity issues is not being addressed at the appropriate level. Therefore, it is recommended that intensive training courses for citizens within Montenegro be organized.

### 3.4 Education, training and awareness in Albania

Compared to other Balkan countries, Internet security is high on the agenda of Albania's institutions. In this regard, Albania's National Security Strategy (2014-2020) classifies cyberattacks as very high risk (of higher importance). Albania, as a NATO member, has also signed a co-operation agreement with the NATO Cyber Response Center (NCIRC) to strengthen internet protection (MoD). According to (GRA), in Albania currently, public and private universities and colleges offer educational courses in cybersecurity-related fields, such as information security, network security, and cryptography, but cyber security-specific courses are not yet offered.

**Table 2:** Evaluation of education, training, and awareness-raising in Western Balkans

	Bosnia	Albania	Kosovo	Montenegro	North Macedonia	Serbia
Address Cybersecurity by the Government	Low	High	Low	Low	Medium	Medium
Cybersecurity Education in primary and secondary schools	Cybersecurity related topics include less than a year of lessons, within existing Information Technology courses	Cybersecurity related topics include less than a year of lessons, within existing Information Technology courses	Cybersecurity related topics include less than a year of lessons, concentrated on no more than one module within existing Information Technology courses	Cybersecurity related topics include less than a year of lessons, within existing Information Technology courses	Cybersecurity related topics include less than a year of lessons, within existing Information Technology courses	Cybersecurity related topics include less than a year of lessons, within existing Information Technology courses
Cybersecurity Education in higher education level	Computer science courses are offered that may have a security component, but no	Some Universities are offering Masters in Information Security and a Professional one-year Masters of Science degree	Only one Private College, (AAB) offers cybersecurity-related courses. The other Universities/Colleges offer	Only one University offers a cybersecurity program at Master Level (1-year program). Other Universities/Colleges offer	There are two undergraduate and master's programs specializing in cybersecurity	Distance learning Masters Programs in Cybersecurity. Also, in Serbia, there are some Universities offering bachelor and

	<b>Bosnia</b>	<b>Albania</b>	<b>Kosovo</b>	<b>Montenegro</b>	<b>North Macedonia</b>	<b>Serbia</b>
	cybersecurity-related courses are offered	program in cybersecurity	courses that have a security component, but not clear cybersecurity programs	courses that have a security component, but not clear cybersecurity programs		master degree in cybersecurity
Cybersecurity Training	Ad-Hoc	Very Limited	Limited	Limited	Very Limited	Limited
Skills	Low	Limited	Low	Low	Low	Limited
Awareness-raising	Low	Limited	Limited	Medium	Limited	Limited

#### **4. Cybersecurity education, training and awareness in Kosovo**

According to the Kosovo Agency of Statistics (Morina, H et al), in 2017 the percentage of households that had access to the internet in Kosovo was 88.8%, in 2018 it was 93.2%, while in 2019 there were no changes in the percentage of household access where 93.2% of households in Kosovo had Internet access at home from any device. This percentage is higher than all Countries in the region but also higher than EU member states. The large use of the Internet in Kosovo means that there is a lot of data online. With the dynamic growth of online data, cyber risks are exponentially increasing for households, but also networks managed by the Government of Kosovo, operators of critical national infrastructure, private institutions as well as financial sectors. Therefore, Kosovo should prioritize Internet security issues, in particular, to develop advanced cybersecurity education programs.

In 2011, Kosovo approved a Strategy for Security and the action plan of this strategy. Later, according to the plan of a strategic document, the security strategy and defense strategy were planned in 2015. Although these strategic documents have been approved, implementation in practice has not been at all satisfactory. Furthermore, no cyber-defense policy or strategy exists; no coordination in response to malicious attacks on military information systems and defense network infrastructure has been established.

In recent years, Kosovo has made considerable progress in raising awareness of technology users about the potential risks in Internet usage, in both; public and private sectors. Many awareness campaigns have been developed at a national level by the National Agency for Personal Data Protection (NAPDP), the 'Privacy and Digital Age Awareness Programme', and some Private Institutions.

The Ministry of Science and Education has placed ICT and security issues as part of the curricula for all levels of education. This is reflected in the efforts to build programs in cybersecurity. In Kosovo, there are some Universities and Colleges offering programs for information-security education and training as well as cybersecurity courses. But, only one private College offers a 3-year Bachelor's Degree in cybersecurity (KAA). Whereas, in 2011 the Kosovo Institute for Public Administration developed some training policies.

Based on the above data, it can be said that so far there have been numerous activities in the development of cybersecurity education strategies. But in a way, these strategies have not been sufficient to raise awareness and educate information technology users to defend themselves against cyberattacks. Some of these strategies are published with no action to date, others have been deficient, and there has been very poor coordination between the various institutions in this sensitive and important field for National Security. The fact, there is only one institution in Kosovo that offers a cybersecurity program shows that not enough has been done in this regard. Another fact is that so far there has been no clear initiative to introduce the basic cybersecurity courses into elementary and high school. Given the best practices of developed countries, this initiative should be the focus of the Government Institutions of the Republic of Kosovo. There is no cooperation between private and public institutions to develop quality education and awareness programs on cybersecurity. Public-private partnership is the key factor for success in this area.

As for cybersecurity training programs, many public and private institutions organize training sessions. Some institutions train staff in the most developed countries where trainers are certified for cybersecurity. Even for training programs, it can be said that not enough has been done so far. Cybersecurity training programs are executed in an ad-hoc manner. Also, there is poor coordination of training programs. These efforts are splintered

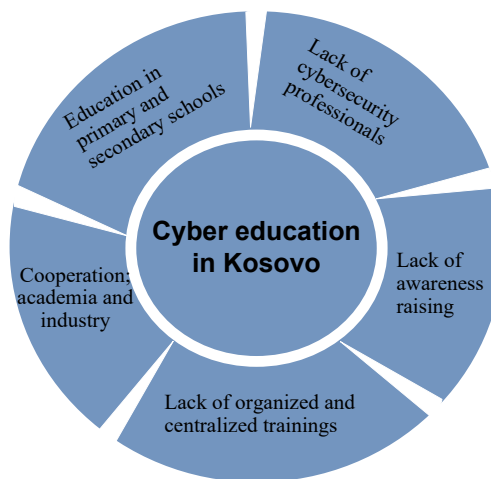


as each Institution is trying to train its staff. A central body should be designated to coordinate the training of IT staff and users of technology. In some institutions, especially public ones, there is a reluctance to spend money on staff training for cybersecurity. These Institutions do not have a unit dedicated solely to cybersecurity; they consider these staff training as an expense, not an investment.

Awareness-raising across society remains at a very low level, except for younger generations. The Ministry of Science and Education, Ministry of Local Government Administration, and National Agency for Personal Data Protection developed an awareness campaign program. The main aim was to increase awareness for the protection of personal data. The target groups are public and private institutions, with a focus on youth, students, etc. Despite the ongoing implementation of the cyber awareness program, we must say that there is no any framework implemented for coordinating and evaluating the effectiveness of such a campaign.

## 5. Identifying the key factors in cybersecurity education in Kosovo

Cybersecurity education in Kosovo depends on some factors affecting universities' decisions to incorporate security content in their curricula and the factors influencing HEIs abilities to implement security instruction. Based on our research study of HEIs in Kosovo, we have addressed some factors according to their relevance. Some of these factors can be found in Table 2.



**Figure 1:** Key factors in cybersecurity education in Kosovo

We have identified 5 main factors in cybersecurity education in Kosovo, (See Figure 1). In the following section, we will describe each of these factors.

**Lack of cybersecurity education in primary and secondary schools in Kosovo:** In primary and secondary Schools, cybersecurity-related topics include less than a year of lessons, concentrated in no more than one module within existing Information Technology courses. This is not enough, at least an introduction to cybersecurity is warranted.

**Lack of cybersecurity professionals:** There are few professionals with formal education in cybersecurity in Kosovo. There are some specialists who not necessarily are teaching cybersecurity because they are teaching other related subjects. Also, we have to add that many professors teaching cybersecurity classes were educated during a time when HEIs did not provide cybersecurity programs or content. There are a few exceptions with some of the professors educated overseas. But, in general, in Kosovo, the number of cybersecurity skilled professionals does not meet the demand.

**Lack of cooperation between academia and industry;** The level of cooperation between academia and industry is very low. There is some interaction regarding topics such as software engineering and networking, however, this cooperation is limited and does not include cybersecurity issues. The HEIs in Kosovo are experiencing difficulties understanding Industry needs in terms of cybersecurity skills.

**Lack of organized and centralized training for cybersecurity:** In Kosovo, there are some incentives for cybersecurity training, while the state budgets for training, research, and development have been allocated. In

general, cybersecurity training programs are executed in Kosovo in an ad-hoc manner (Bada, M). The private sector is better organized in terms of cybersecurity training and they do follow risk-management policies. However, the private ICT Companies still lack specific cybersecurity training. AAB College, which is the only College in Kosovo offering a 3-year cybersecurity program, established a Cybersecurity Center in March 2020. The Cyber Center will be engaged in training and research activities in cybersecurity.

**Lack of awareness-raising on cybersecurity:** There is increasing awareness of cyber risks in both; public and private sectors. Society, in general, is characterized by a feeling of fear of cyber threats because of a lack of understanding of the benefits and risks of the digital World. According to surveys conducted, private companies are more aware of cybersecurity than public institutions (Bada, M). In general, awareness-raising in Kosovo remains at a low level.

## **6. Conclusions and recommendations for cybersecurity education in Kosovo**

Technology, processes, and people are equally important to ensure success in cybersecurity programs. Even the implementation of more advanced technology would not succeed if there was no awareness, education, and training of users. Awareness-raising, education, training, principles, policies, processes, and programs are essential components of any cybersecurity strategy.

No single entity or group of stakeholders can secure cyberspace alone. Everyone's commitment is needed, including governments, organizations of all sizes as well as consumers to secure their systems. In this regard, the biggest impact is education and awareness-raising.

Currently, in Kosovo, there is little formal education in cybersecurity addressing students at the university level. As we mentioned, there is only one accredited 3 years program, which offers cybersecurity classes. The main reason for this situation is the lack of a clear strategy in cybersecurity education. Cybersecurity includes not only technical issues, legal and human issues as well. So, we are dealing with a very complex and specific field that is also considered multidisciplinary, thus it is impossible to create experts that can cover all parts of cybersecurity.

First, we have reviewed the objectives for some of the EU Countries. According to the official reports and strategies for the reviewed Countries, cybersecurity education is a common objective for all of the EU countries.

Second, with scientific research and official reports on cybersecurity maturity level, it can be concluded that these Countries should and can do much more in educating, training, and raising awareness of their citizens. We have considered some important factors in each of these countries, such as education in primary and secondary schools, education in higher education level, training, skills, and awareness-raising on cybersecurity. From the analysis, it can be concluded that all Western Balkan countries have a low level of education, training, awareness, and skills specific to cybersecurity. Kosovo, as the newest state within the Western Balkans, also lags. Lately, some initiatives and strategies have been noted in Kosovo, but most have not been implemented in practice, or have been uncoordinated and ineffective. In general, we have identified 5 main factors affecting cybersecurity education in Kosovo. Included are lack of cybersecurity education in primary and secondary schools, lack of cybersecurity professionals, lack of cooperation between academia and industry, lack of organized and centralized training for cybersecurity, and lack of awareness-raising on cybersecurity. Therefore, our recommendations can be summarized as follows.

In the short-term, the government and responsible institutions should urgently focus their efforts on the current situation. Short-term objectives should focus on higher education curricula. Specifically, we recommend that the Government should continue with the following activities; creating cybersecurity curricula's not only for higher education but also for K-12. Knowing the importance of training programs, Government should develop more professional training programs and coordinate the training process from a centralized body. Regarding the awareness campaign, the Institutions should coordinate this process as well. Currently, there are many initiatives, but they are organized in an ad-hoc manner.

In the long-term, Kosovo should establish a cybersecurity education system at the national level. The Government should create a research and development environment in the cybersecurity field. They should also be focused on strengthening the training and education system through a public-private partnership, provide an incentive for the private sector to invest in the security field (internships and scholarships), and organize

cybersecurity exercises. Last but not least, the government should find ways to increase international cooperation in this field.

## **Acknowledgements**

The Fulbright Scholarship Program, Capitol Technology University and AAB College Kosovo made the results of this collaboration possible.

## **References**

- ASK. Morina, Hydai, Ukaj, Marigona, Cakolli, Ahmet, 2019. Kosovo Agency of Statistics, ISBN 978-9951-22-616-5, 2019, Survey Results regarding the usage of Information Technology and Communication. [ONLINE] Available at: <https://ask.rks-gov.net/>. [Accessed 18 February 2020].
- Bada, M., 2015. Cybersecurity Capacity Assessment of the Republic of Kosovo. *Global Cyber Security Capacity Centre*.
- Bush, G.W., 2003. The national strategy to secure cyberspace. *The White House, Washington*.
- DHS, United States. Office of Homeland Security, 2002. National strategy for homeland security (No. 87). Office of Homeland Security.
- Dlamini, I.Z., Taute, B. and Radebe, J., 2011. *Framework for an African policy towards creating cybersecurity awareness*.
- EDUCAUSE. 2019. Don Welch, Creating a Cybersecurity Strategy for Higher Education. [ONLINE] Available at: <https://er.educause.edu/articles/2019/5/creating-a-cybersecurity-strategy-for-higher-education>. [Accessed 14 September 2020].
- ENISA Austria. 2013. Austrian National Cyber Security Strategy. [ONLINE] Available at: <https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map?selected=Austria>. [Accessed 5 March 2020].
- ENISA Croatia. 2015. Croatian National Cyber Security Strategy. [ONLINE] Available at: <https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map?selected=Croatia>. [Accessed 1 March 2020].
- ENISA Finland. 2013. Finnish National Cyber Security Strategy. [ONLINE] Available at: <https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map?selected=Finland>. [Accessed 1 March 2020].
- ENISA France. 2015. France National Cyber Security Strategy. [ONLINE] Available at: <https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map?selected=France>. [Accessed 5 March 2020].
- ENISA Hungary. 2018. Hungarian National Cyber Security Strategy. [ONLINE] Available at: <https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map?selected=Hungary>. [Accessed 1 March 2020].
- ENISA Italy. 2017. Italian Cybersecurity Action Plan. [ONLINE] Available at: <https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map?selected=Italy>. [Accessed 8 October 2020].
- ENISA NCSS. 2012. National Cyber Security Strategies. [ONLINE] Available at: <https://www.enisa.europa.eu/publications/cyber-security-strategies-paper>. [Accessed 14 October 2020].
- ENISA Slovakia. 2015. Slovakia National Cyber Security Strategy. [ONLINE] Available at: <https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map?selected=Slovakia>. [Accessed 1 March 2020].
- ENISA Slovenia. 2016. Slovenian National Cyber Security Strategy. [ONLINE] Available at: <https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map?selected=Slovenia>. [Accessed 1 March 2020].
- ENISA. 2015. Cybersecurity Education snapshot for workforce development in the EU. [ONLINE] Available at: <https://resilience.enisa.europa.eu/nis-platform/shared-documents/wg3-documents/cybersecurity-education-snapshot-for-workforce-development-in-the-eu/view>. [Accessed 12 November 2020].
- EU Cybersecurity. 2015. EU Cybersecurity Maturity Dashboard - A path to secure European Cyberspace. [ONLINE] Available at: <http://cybersecurity.bsa.org/>. [Accessed 6 March 2020].
- GFBH, Government of Bosnia and Herzegovina. 2019. Cybersecurity capacity review Bosnia and Herzegovina. [ONLINE] Available at: <http://mkt.gov.ba/>. [Accessed 4 March 2020].
- GRA, Government of Albania. 2019. Report on cybersecurity maturity level in Albania. [ONLINE] Available at: <https://cesk.gov.al/>. [Accessed 5 February 2020].
- GRNM, Government of Northern Macedonia. 2018. Cybersecurity capacity review Former Yugoslav Republic of Macedonia. [ONLINE] Available at: <http://mioa.gov.mk/>. [Accessed 5 February 2020].
- GRS, Government of Serbia. 2017. Strategy for the Development of Information Security in the Republic of Serbia for the period from 2017 to 2020. [ONLINE] Available at: <https://ials.sas.ac.uk/eagle-i/official-gazette-republic-serbia>. [Accessed 5 February 2020].
- ITU. 2015. Global Cybersecurity Index & Cyberwellness Profiles. [ONLINE] Available at: [https://www.itu.int/pub/D-STR-SECU-2015#:~:text=The%20Global%20Cybersecurity%20Index%20\(GCI,the%20forefront%20of%20national%20plans.](https://www.itu.int/pub/D-STR-SECU-2015#:~:text=The%20Global%20Cybersecurity%20Index%20(GCI,the%20forefront%20of%20national%20plans.) [Accessed 5 March 2020].

**Arianit Maraj, Cynthia Sutherland and William Butler**

- KAA, Kosovo Accreditation Agency. 2019. Study programmes at Kosovo higher Education. [ONLINE] Available at: <http://www.akreditimi-ks.org/new/index.php/en/>. [Accessed 18 February 2020].
- KCSS. 2014. Cyber security winter School. [ONLINE] Available at: <http://www.qkss.org/>. [Accessed 17 April 2020].
- Maraj, A., Jakupi, G., Rogova, E. and Grajqevci, X., 2017, June. Testing of network security systems through DoS attacks. In *2017 6th Mediterranean Conference on Embedded Computing (MECO)* (pp. 1-6). IEEE.
- Maraj, A., Rogova, E. and Jakupi, G., 2020. Testing of network security systems through DoS, SQL injection, reverse TCP and social engineering attacks. *International Journal of Grid and Utility Computing*, 11(1), pp.115-133.
- Maraj, A., Rogova, E., Jakupi, G. and Grajqevci, X., 2017, October. Testing techniques and analysis of SQL injection attacks. In *2017 2nd International Conference on Knowledge Engineering and Applications (ICKEA)* (pp. 55-59). IEEE.
- Minović, A., Abusara, A., Begaj, E., Erceg, V., Tasevski, P., Radunović, V., Klopfer, F. and DiploFoundation, G., 2016. *Cybersecurity in the Western Balkans: Policy gaps and cooperation opportunities*. Research Report, DiploFoundation, Geneva, Switzerland, 2016. Accessed January 12, 2017. <https://www.Diplomacy.edu/sites/default/files/Cybersecurity%20in%20Western%20Balkans.pdf>.
- MoD, Ministry of defense Albania. 2014. Strategy for cyber security protection. [ONLINE] Available at: <http://www.mod.gov.al>. [Accessed 5 February 2020].
- Poposka, V., 2016. THE URGE FOR COMPREHENSIVE CYBER SECURITY STRATEGIES IN THE WESTERN BALKANS. *Information & Security*, 34(1), pp.25-36.
- Rizmal I., Guide through information security in the Republic of Serbia 2.0, Publishers: OSCE Mission to Serbia, Belgrade, Unicom Telecom, Belgrade, IBM, Belgrade, Juniper, Belgrade, 2018, ISBN 978-86-6383-078-3
- Šendelj, R. and Ognjanović, I., 2015. Cyber Security Education in Montenegro: current trends, challenges and open perspectives. In *The 7th annual International Conference on Education and New Learning Technologies (EDULEARN15)*.
- Von Solms, S. and von Solms, R., 2014. Towards Cyber Safety Education in Primary Schools in Africa. In *HAIISA* (pp. 185-197).

# Studying the Challenges and Factors Encouraging Girls in Cybersecurity: A Case Study

Arianit Maraj<sup>1</sup>, Cynthia Sutherland<sup>2</sup> and William Butler<sup>2</sup>

<sup>1</sup>AAB College, Faculty of Computer Sciences, Kosovo

<sup>2</sup>Capitol Technology University, Laurel, Maryland, USA

[arianit.maraj@universitetiaab.com](mailto:arianit.maraj@universitetiaab.com)

[cevalentine@captechu.edu](mailto:cevalentine@captechu.edu)

[whbutler@captechu.edu](mailto:whbutler@captechu.edu)

DOI: 10.34190/EWS.21.004

**Abstract:** Today, there is a clear gender gap in cyberspace. Many barriers hinder the advancement of women in cybersecurity. In addition, a lot of studies point out that IT (Information Technology) is a more male-oriented world, which is one of the reasons why so few women pursue a career in cybersecurity. In this paper, we will try to explore the challenges and possible reasons for and suggest solutions to address this gap. Encouraging girls in cybersecurity will not only contribute for developing a strong cybersecurity workforce, but also one that should be prepared for offering more cybersecurity solutions. Through this paper, different stakeholders will also better understand their roles, duties, responsibilities and benefits for reducing a gender gap. The only way to cope with technological challenges is through educating and raising awareness of young women. Our idea is to provide a roadmap to cybersecurity awareness and training of young girls in developing Countries, with a focus in Western Balkans Countries, and help them to understand that the cybersecurity field offers a rewarding career for women. Our thesis is that increased awareness opportunities offered would help to increase the participation of girls and women in STEM (Science, Technology, Engineering and Mathematics) for the long term. In general, this paper will identify the challenges in encouraging young girls to seek cybersecurity related careers.

**Keywords:** cybersecurity, STEM, girls, education, mentoring, training

---

## 1. Introduction

Cybersecurity is a mission essential function for all businesses. Due to the Internet, every organization either possesses or has the potential to having a digital presence in Cyberspace. As more businesses, organizations, and government institutions migrate to using digital data and technology, they require an educated and well-trained staff to protect their digital data and technology. Increased dependency on the Internet has increased the rise of cyber-attacks at an alarming pace. Application of traditional criminology theories to Cyberspace highlights an increase in capable guardians to protect vulnerable assets can lead to a reduction in cyberattacks. However, there is a global shortage in numbers and skills of a capable cyber workforce to address these requirements. The professional skills shortage could be overcome by addressing the gender imbalance, which exists throughout the World. According to (Reed, J et al), in 2017 only 11% of global security workforce were female. In Europe, this was as low as 7%. Whereas in US, according to (LeClair, J et al) there was a 10-15% representation of women in cybersecurity jobs. Nevertheless, at the end of 2019, women represent more than 20% of the global cybersecurity workforce. It can be seen that in recent years there has been an increasing trend of women being involved in cybersecurity. However, this percentage is very small, so there is a need to do much more in this regard. This percentage in developing Countries is much smaller, so it is imperative to take measures to support women in cybersecurity, especially in developing Countries.

The biggest challenge today for IT security decision makers is a lack of cybersecurity staff (25%) and lack of staff with the right cybersecurity skills (22%) (Higgins, K.). However, organizations can adopt cybersecurity best practices to help hack the cybersecurity talent gap by recruiting more women in cyberspace. Women make up half of total IT users, though they are largely underrepresented in the technology business world. In this new era, it is time for women to become actively involved in the technology world.

By encouraging women to be involved in technology, we will be able to convince upcoming generations that there is just as much room for women as men (MEST). By 2021, there is expected to be 3.5 million unfilled cybersecurity roles worldwide and the idea is to challenge the stereotype that the industry is male-dominated (USEJOURNAL). The unfilled cybersecurity jobs should not be considered as personnel issues or percentage improvement; they should be considered a national security issue and women can help us solve the problem quickly.

In developing countries such as Kosovo women's representation in technology is very low; therefore, representation in cybersecurity is extremely low. It is worth mentioning that even academic institutions in Kosovo have not paid much attention to introducing cybersecurity programs to young women. In Kosovo, to our knowledge, to date there is only one bachelors program in cybersecurity. At this college, the number of girls who have registered for this course is symbolic. Therefore, there is an urgent need to raise the awareness of women to pursue cybersecurity as a career.

Given the importance of education and certification, women with cybersecurity education and skills should have a clear path to career advancement to gain the right qualifications to assume leadership roles. With this leadership comes more responsibility and credibility, as well as a boost in salary. These gains are important not only for current women in cybersecurity, but also for future generations in Kosovo.

The main goal of this manuscript is to raise the awareness of Kosovo High School women about cybersecurity as a career and to identify the main factors encouraging girls in cybersecurity. This paper offers insights into the current representation of women in the STEM field, especially in cybersecurity in Kosovo and in Western Balkans (WB) countries. Even in EU Countries, men dominate in the ICT Field. Men accounted for at least 8 out of every 10 ICT specialist in the majority of EU member States. According to Eurostat, in Bulgaria women accounted for 28.3% of IT specialist in 2018, which is the highest among EU members. In Lithuania and Romania the percentage of Women in ICT was around 25% in 2018, whereas in Estonia, Sweden and Finland this percentage was 20%. This percentage in WB Countries is lower. Currently, statistics for women in science in WB Countries are a bit more positive, but work still needs to be done for improving the percentage of women in Technology.

In the next session of this paper, we will describe the literature review. In section 3, we will describe the main challenges to engage women in cybersecurity. In section 4, we will present the current state of women involvement in cybersecurity in WB. The main challenges and factors encouraging women in cybersecurity in Kosovo are presented in section 5. Conclusions and recommendations are drawn in the last section.

## **2. Literature review**

In this section, a literature review will be presented in order to provide a synthesis of relevant literature concerning the factors encouraging women in cybersecurity field.

In (Peacock, D., et al) the authors consider the impact of gender in the global cybersecurity industry. They developed a survey in order to determine the gender gap in cybersecurity field. According to this survey, they found out that there is a need to encourage girls and young women at school, to further their education and to study the cybersecurity discipline.

Authors in (Rowland, P, et al) discuss the CybHER model for engaging and supporting young women in cybersecurity while anchoring them to this field. By providing 5 different interventions, CybHER seeks to empower, motivate, educate, and anchor girls to cybersecurity.

Social connections beyond the classroom, have been found to be an effective measure to retain female students in cyber degrees (Cohoon, J.P). Mentorship and socialization are very important factors to address some deficiencies found in traditional approaches to cybersecurity education. These factors were studied in (Seymour, E). The positive effects of socialization using strategies in teaching cybersecurity to undergraduate women have been shown in (Flushman, T, et al) as well.

According to a survey performed in ((ISC)<sup>2</sup>), Cybersecurity professionals are more than twice as likely to be male. In this survey, 30% of respondents were women. Among respondents with security-specific titles, 23% of study participants were women. The highest percentage of women cybersecurity professionals came from LATAM (39%) and North America (34%).

Authors in (Bear, J.B. et al) evaluated the gender gap in STEM and the effects of gender balance in teams. Based on their survey, they found that gender diversity could also enhance group processes, which are increasingly important in the production of science. The psychological factors related to gender differences are investigated in (Bashir, M et al). The researchers investigated the challenges, barriers, skills and knowledge required for women in cyberspace to view their success throughout their careers.

### **3. The challenges to involve girls in cybersecurity**

We are facing an acute shortage of cybersecurity professionals. Government and non-government institutions worldwide are concerned about a cybersecurity skills crisis (ISACA). According to (Morgan, S), by 2021 there will be 3.5 million unfilled cybersecurity jobs globally, in comparison from 1 million positions in 2014.

In general, girls are widely underrepresented in STEM related subjects and to include cybersecurity. Although adolescent girls are involved in other STEM (science, technology, engineering and mathematics) related subjects, they are less likely than boys to pursue computer science (Jethwani, M.M. et al).

According to (Schurr A.), the lack of women in IT and especially in cybersecurity represents a failure to capitalize the benefits of diverse opportunities and the diversity can bring the brightest problem-solvers to the table.

In order to increase the numbers of girls in cyberspace, the focus should be on engaging them during adolescence. However, there are factors and challenges that discourage girls from entering in cyber related professions, such as: lack of awareness, lack of support from their families, discrimination, culture (male dominated), etc. Therefore, there is a need to follow promising practices that will contribute to the improvement of cybersecurity interest of girls. The underrepresentation of girls in computer sciences is indicative of some unique challenges that girls face. Studies found that usually, girls in computer science classes express less confidence than boys (Cooper, J., Moorman, P. et al). Master et, al. (2015) found that high school girls experience lower feelings of belonging in computer science courses when compared with boys. However, if we provide them an educational environment that does not fit current science stereotypes will increase their interest in computer sciences courses and in general could help reduce the gender disparities in computer science enrolment. Master, A., et al (2016), showed that teenage girls lack awareness regarding the opportunities that a cybersecurity career offers. Another challenge is that girls are often not supported by their families to pursue their dreams in STEM. In addition, another challenge is finding role models and mentors to help girls become aware of the benefits of careers in cybersecurity. This challenge is very pronounced in developing countries, such as Kosovo. Apparently, low confidence, male perceptions of the field, and limited access to role models in Kosovo impacts girls' involvement in cybersecurity.

### **4. Background; current state of women's involvement in cyber space in Western Balkans (WB) countries**

Twenty years after the end of armed conflicts in WB, the Region remains fragile. There is a high potential for conflict between some countries in the region. Corruption, lack of media freedom, and lack of a defined cybersecurity strategy are the main drivers of regional instability. The WB region has been a synonym of political and economic instability for a long period of time. Like the rest of the world, regional economies have faced numerous security threats and challenges, with an increasing number of attacks occurring in cyberspace (ITU). High internet access within the region means more threats from the Internet seriously jeopardizing the security of the user data.

As the whole world faces the risk of cyber-attacks, the governments of the WB have begun to list this threat as one of their national priorities. One of the biggest challenges to strengthening national cybersecurity capabilities for developing countries is identifying strengths and weaknesses, threats and opportunities. This can be achieved by training and educating the younger generations to give them the opportunity to protect sensitive data on the Internet. The only way to respond to the demands of the industry for cybersecurity professionals is with greater female involvement. So far, the percentage of women involved is very low, especially in developing countries. Therefore, we consider that diversity can provide benefits to Industry, especially in cybersecurity.

Most of the WB Countries have created national cybersecurity strategies. But, what is noticeable is that these end-to-end strategies have not achieved full implementation. Also, in all WB countries there is a significant lack of professional staff, especially implementing plans to increase the number of women in cybersecurity. Almost none of the WB have clearly foreseen the strategy of involving women in this sector. In addition, educational programs have recently been developed to meet the demand for online security professionals in the current and future job market. However, there is still a lot of work to be done. Overall, the WB region is undergoing a process of transforming awareness of the dangers and opportunities of cybersecurity. Table 1 analyses some of the key factors that influence women's encouragement of cybersecurity. As can be seen from the Table 1, the WB Countries are almost at the same level in terms of organization and activities undertaken for the engagement of

women in cybersecurity. In addition to awareness raising, these Countries are at a satisfactory level. In all other initiatives, there is a low degree of organization and the activities undertaken to improve the representation of women in cybersecurity.

**Table 1:** Main factors that influence women in cybersecurity field

	Women's Cyber forums and seminars	Framework for professional training and	Mentoring programs	Resource pipeline from High School to college to the	Awareness rising	Family support	Society support	Incentives for women-sponsoring cybersecurity	Connect women with cybersecurity early on
<b>Kosovo</b>	Low	Ad-hoc	No	No	Medium	Low	Low	Low	Ad-Hoc initiatives
<b>Albania</b>	Low	Ad-Hoc	No	No	Medium	Low	Low	Low	Ad-Hoc initiatives
<b>North Macedonia</b>	Medium	Ad-Hoc	No	No	Medium	Low	Low	Low	Ad-Hoc initiatives
<b>Montenegro</b>	Low	Ad-Hoc	No	No	Medium	Low	Low	Low	Ad-Hoc initiatives
<b>Serbia</b>	Medium	Ad-Hoc	No	No	Medium	Low	Low	Low	Ad-Hoc initiatives
<b>Bosnia and Hercegovina</b>	Low	Ad-Hoc	No	No	Medium	Low	Low	Low	Ad-Hoc initiatives

**Montenegro:** In the cybersecurity strategy document in Montenegro, there is no mention of women's involvement in technology. This shows best that there is no state strategy that would increase the percentage of women's participation in cybersecurity. However, in Montenegro, during last 5 years, there have been a lot of activities regarding cybersecurity. The following activities were held: Safe Internet Day, debates and presentations for children, teachers and parents in primary and secondary schools were held in the framework of some projects. But the most important achievements towards gender equality and empowerment of women over the past 5 years are considered to be work on the elimination of gender stereotypes, improving status of women in the security sector, promoting jobs in the security sector within younger population, with a special focus on girls and women, and the professional empowerment of women in the security sector (MHMR).

**Albania:** In Albania, it has been recognized that there is an increasing number of women contributing to cybersecurity, breaking gender stereotypes about cyberspace. Currently, there are organized events bringing together a lot of women sharing success stories in cyberspace and inspiring young women to work the cybersecurity. There have been also some other collaborations with different Government entities, led by women, to enforce cybersecurity law. However, there is a huge gap between the growing demands of the private sector for cybersecurity professionals and the programs that training and education institutions provide to young people. Since 2017, conferences on data security and privacy have been organized in Albania. Some of the conferences were organized jointly with Kosovo authorities and were held in Albania and Kosovo.

**North Macedonia:** North Macedonia is one of the WB states that demonstrates how fast progress can be achieved towards national cybersecurity capacity building. In 2018, the government adapted the National Cybersecurity Strategy 2018-2022. They also hosted the WB Digital Summit, the first ever summit of this kind in the country. The main focus of this summit was cybersecurity. In 2019, North Macedonia hosted the ITU Cybersecurity Workshop with the goal of facilitating regional and international cooperation in this field. In North Macedonia, as in other WB countries, there is a great shortage of professional staff in cybersecurity, especially women in cybersecurity. So far, not enough has been done to encourage young women to become more involved in this profession. Even in the cybersecurity strategy, attention has not been paid to the treatment of the gender gap in the field of cybersecurity.

**Kosovo:** Kosovo has adopted a strategy for cybersecurity since 2011. Although these strategic documents have been approved, implementation in practice has not been satisfactory. Furthermore, no cyber-defence policy or strategy exists; no coordination in response to malicious attacks on military information systems and defence network infrastructure has been established. In recent years, Kosovo has made considerable progress in raising the awareness of technology users to the potential risks of Internet usage. Many awareness campaigns have been developed at a national level. Since 2017, conferences on cybersecurity have been organized in Kosovo.



Recently, private sector organizations have taken several initiatives to support and motivate women to attract to cybersecurity. It is worth mentioning that a cyber-academy is offering some scholarships for girls and women in a program called "cyber her". Other events related to cybersecurity, conferences, raising awareness, etc., have been organized, but in ad-hoc manner. However, in Kosovo there is no any event and support for mentoring programs for girls, strategy for pipelining girls from high school to the college, or framework for professional training, education framework and curricula's, etc.

**Serbia:** The cybersecurity strategy adopted in 2017 clearly defines priority areas that include the security of information systems. This strategy is considered a significant move forward in efforts aimed at building capacity in information security of the society as a whole - from elementary school to study programs at universities (GRS 2017, OSCE 2018). The Diplo Centre with the support of the OSCE Mission to Serbia, organized some workshops for representatives of all the key institutions, private sector, and civil society dedicated to discussions on cybersecurity. These workshops also focused on the development of a national strategic framework for cybersecurity (CEAS OSCE, Rizmal I., et al). In addition to these events, a large number of conferences and public discussions were organized on specific aspects of this field. In 2019, Serbia co-organized with the ITU a Belgrade Women's forum. Despite the fact that in general, Serbia has made progress in terms of professional advancement of staff, there remains a small percentage of women's commitment to cyberspace. In particular, there is no real organization and support for mentoring programs for girls, resources pipeline from high school to the college, or other initiatives such as connecting women with cybersecurity early on.

In general, the percentage of women's participation in cybersecurity in WB is still so all responsible entities should look at these figures with concern and commit to increase this percentage in the coming years.

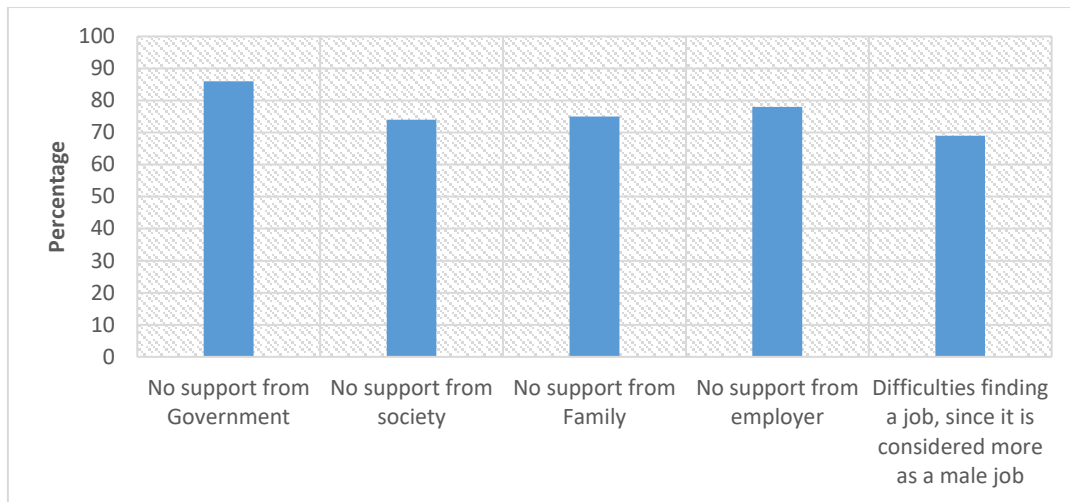
There are several studies stating the importance of cyber education in WB (OSCE, Rizmal I., Maraj, A. et al, Minović, A. et al, Šendelj, R. et al, Poposka, V. et al). A review of major projects and funding opportunities in cybersecurity in the WB by major international organizations is presented afterwards. The cybersecurity situation at the WB is consistent for all Countries. There are also a lot of initiatives for empowering girls in cybersecurity. Nevertheless, there is still much work to be done.

## **5. The challenges and factors encouraging women in cybersecurity in Kosovo**

According to Reed, J et al (2017), girls and young women historically were more likely to be excluded from education. There are a lot of problems related to education, such as: funding, strategy and not benefiting from real experts among decision makers. The main challenge in developing countries is providing the opportunities for students, which are underrepresented and broaden that pipeline, especially in STEM, where many rural and urban teachers do not have laboratory equipment. Today, more female students attend school in Kosovo than ever before, but the overall quality of education is poor, especially in rural schools. There are numerous factors which affect rural girls, such as: lack of support from their family and society, lack of good teachers in STEM subjects, and in STEM careers as well as poorly developed infrastructure in fixed Internet.

In general, only one in eight women are formally employed in Kosovo (World Bank Report 2015). This disparity is more pronounced in ICT and especially in rural areas. Based on a survey conducted with only the category of women with samples taken from all over Kosovo (Farnsworth, N. et al and STIKK 2014), it was noted that a high percentage of them stated that they did not have any support from the government (86%), family (75%), society (74%), employers (78), but they have also had difficulties finding a job in ICT field since it is considered a male job (69%), Figure 1. Based on this survey (Farnsworth, N. et al), the percentage of women working in ICT is 22%. This percentage is smaller in the field of cybersecurity, about 13%. So far, to our knowledge, there is no survey that identifies the main challenges women face in engaging in cybersecurity. However, we can say that the challenges are the same for women in the field of ICT.

The field of ICT has been developed recently in Kosovo. However, with all the recent advances, remains far behind in this regard. Due to the low number of women involved in ICT field, several projects have been established in Kosovo to raise the awareness and interest of girls and women to get involved in the ICT. One of these projects is "Girls in ICT", a project that has been implemented by NGO KS-Kosovo. The main purpose of this project was to encourage high school girls to use ICT and also to inform them about the possibilities of career opportunities in the ICT sector.



**Figure 1:** Challenges faced by women in the ICT field in Kosovo (Farnsworth, N. et al)

Another initiative is Girls Coding Kosovo-GCK, which aims to encourage and motivate women to choose the programming field as an opportunity for their profession and to encourage them to code, launch companies or work as other exciting startups in the country.

Prejudice against women in the field of ICT and the mentality that the profession in this field belongs to males are just some of the barriers that women face in Kosovo. According to Farnsworth, N. et al (2018), there are company managers who have stated that they will not employ women in their companies because of maternity leave and benefits this law provides to women. There are many women in Kosovo who have begun to choose this path of education, but in the Kosovar market there is still a lack of employment in the field in which they are studying. It has also been noted that most of those who have chosen to study in the ICT field belong to the male gender, because there is still a stereotype that this field belongs to the male gender only. Another factor contributing to the low participation of women in the workforce is the responsibility of providing care, mainly child care; 32% of women say this hinders their employment (Kosovo Live, 2016).

Women's participation in cybersecurity is extremely low. It is worth mentioning that even Academic Institutions in Kosovo have not paid much attention to advancing technology programs and introducing cybersecurity programs to women. Therefore, there is an immediate need to implement projects and other initiatives in order to raise the awareness of Kosovo High School women about cybersecurity as a career.

To ensure a sustainable future for Kosovo, the Country needs to consider STEM and cyber in particular, as an important catalyst. Kosovo must strengthen girls' education in cybersecurity, as it will not only help increase their participation in cyber activities, but also will help to strengthen the National and Global security.

## 6. Conclusions and recommendations

The rapid development of the cybersecurity field is one of the areas with the fastest growth of opportunities for professionals seeking to improve their careers. Cybersecurity matters for everyone from governments, large companies to small businesses, employees, and even individuals at home. The number of cyberattacks is increasing daily. This means that it is even more important to keep software up to date, hardware protected, and data safe. There exists a shortage of professionals, which have the skills to protect these cyber assets.

We are engaged in a global cyber war and our enemies know that we are not sufficiently prepared to win this war. Our unpreparedness is due to a severe shortage in technical talent. Our best-untapped option is to train and recruit more girls into the technology and cybersecurity fields. We are doing a poor job of marketing and selling these work roles to young women. We need to educate youth - especially women - to the fact that cybersecurity jobs cover wide and diverse positions. Today, working as a cybersecurity expert is a very desirable job, while in the media is still described as strange and complex. We need to change this view, because negative stereotypes and miscommunication is hampering the recruiting effort. It is important to clarify why there is a need to increase women's cybersecurity. It's not just about increasing numbers, but the main argument is that having more women in the workplace has a positive effect on business and National security. We need diversity

because the people we follow, (threat actors, hackers, 'bad guys') also have a variety of different backgrounds and experiences. In addition, the demand for hard-working security professionals is so high and involving more women in cybersecurity would have great impact in filling these empty positions. However, cybersecurity still has some perception problems. Women often don't see technology or cybersecurity as a career, because they are considered to be male professions. These problems are difficult to solve because they are very complex in nature; issues related to culture and education. These problems are even more pronounced in developing countries, such as Kosovo.

## **7. Recommendations**

*Establishing a resource pipeline from High School to college to the Workforce:* There is a need to establish a resource pipeline from High School to College to the Workforce, called CybHER Pathways. The focus of CybHER Pathways should encourage high schools to provide a pipeline for girls or women to enter the cybersecurity profession. This program should identify and target barriers to women's participation in cybersecurity education and careers. This could be achieved by providing well-structured training and awareness opportunities with the main aim of increased participation of women in cybersecurity.

*Cybersecurity training, workshops and seminars:* Basic cybersecurity training should be provided and students should be provided the latest literature in this field. The main goal of the HEIs should be to provide students with a basic knowledge of cybersecurity, but also to make women aware of the benefits of that choice in professions. Organizing workshops and seminars focusing on topics such as Cyber Ethics, CybHER Safety, and Cyber Hygiene are more than necessary in Kosovo. The idea of the workshops should have a dual purpose; learning protection concepts online, and reducing the risks that women are exposed to online threats.

*Education and certification:* Given the importance of education and certification, women in cybersecurity will have a clear path to career advancement and gain the right qualifications for leadership roles. With this leadership comes more responsibility and credibility, as well as a boost in salary. These gains are important not only for current women in cybersecurity but also for future generations in Kosovo.

*Mentoring:* The young women should be provided the opportunity to connect with a mentor in the cybersecurity field from organizations such as Women in Cybersecurity, Society of Women Engineers and other notable women's organizations. The young women should also receive the opportunity to continue their training with packaged training from websites such as girl's who code (<https://girlswhocode.com/>), Cybrary ([www.cybrary.it](http://www.cybrary.it)), Codecademy ([www.codecademy.com](http://www.codecademy.com)), and Khan Academy ([www.khanacademy.com](http://www.khanacademy.com)).

*Additional funding for involving more women in cybersecurity:* From the Government and other responsible agencies, additional funding should be sought to address the lack of computers and Internet for young girls though women oriented initiatives such as the Bill and Melinda Gates Foundation and the Clinton Initiative, and industry charitable contributions such as Barbie/Mattel.

*Promoting cybersecurity as a professional opportunity for girls and women:* This could include capacity building and training. There are a lot of challenges faced by women in cybersecurity sector, such as: no support from government, difficulties in finding job, no enough support from society and family, no flexibility when becoming a mother, no support from employees, etc. All of these challenges exist in developing countries, such Kosovo is. Overcoming these problems can only be done with the implementation of clear policies by the government and with the support of business entities. Specifically, the government needs to focus on and prioritize educating young girls and promoting the cybersecurity profession as an attractive profession for women. The government should initiate incentives for women from the earliest years, to train teachers to encourage girls to pursue cybersecurity careers, to develop curricula that are gender-sensitive, to foster cultural change at cybersecurity events and to mentor girls and young women.

Kosovo has made great progress in closing its gender gap, particularly in education and technology. However, statistics show that sizeable gaps still remain in terms of cybersecurity field. There is a need to consider the role of private sector, academia and the entire society. It is also important to adopt best practices from neighbour countries and educate young women about the growth of cybersecurity and workforce demands. To facilitate awareness raising Western Balkans Countries should undertake joint efforts. Until now, these efforts are limited to Kosovo and Albania. These initiatives should include other Countries within the region as well. Awareness

raising for cybersecurity and promoting equal gender participation through education and sharing of best practices is an important element in the national cybersecurity capacity building.

## **Acknowledgements**

The Fulbright Scholarship Program, Capitol Technology University and AAB College Kosovo made the results of this collaboration possible.

## **References**

- (ISC)<sup>2</sup>. 2019. Strategies for Building and Growing Strong Cybersecurity Teams. [ONLINE] Available at: <https://www.isc2.org/-/media/ISC2/Research/2019-Cybersecurity-Workforce-Study/ISC2-Cybersecurity-Workforce-Study-2019.ashx?la=en&hash=1827084508A24DD75C60655E243EAC59ECDD4482>. [Accessed 25 April 2020].
- Bashir, M., Lambert, A., Wee, J.M.C. and Guo, B., 2015. An examination of the vocational and psychological characteristics of cybersecurity competition participants. In 2015 {USENIX} Summit on Gaming, Games, and Gamification in Security Education (3GSE 15).
- Bear, J.B. and Woolley, A.W., 2011. The role of gender in team collaboration and performance. *Interdisciplinary science reviews*, 36(2), pp.146-153.
- CEAS OSCE, Rizmal I., Radunović V., Krivokapić Đ., Guide through information security in the Republic of Serbia, Publishers: Centre for Euro-Atlantic Studies – CEAS OSCE Mission to Serbia
- Cohoon, J.P., 2007, March. An introductory course format for promoting diversity and retention. In Proceedings of the 38th SIGCSE technical symposium on Computer science education (pp. 395-399).
- Cooper, J., 2006. The digital divide: The special case of gender. *Journal of Computer Assisted Learning*, 22(5), pp.320-334.
- Eurostat. 2018. Proportion of ICT specialists in total employment. [ONLINE] Available at: [https://ec.europa.eu/eurostat/statistics-explained/index.php/ICT\\_specialists\\_in\\_employment#ICT\\_specialists\\_by\\_sex](https://ec.europa.eu/eurostat/statistics-explained/index.php/ICT_specialists_in_employment#ICT_specialists_by_sex). [Accessed 15 April 2020].
- Farnsworth, N., Morina, D., Ryan, D.J., Rrahmani, G. and Robinson-Conlon, V., 2018. and Iliriana Banjska for the Kosovo Women's Network.
- Flushman, T., Gondree, M. and Peterson, Z.N., 2015. This is not a game: early observations on using alternate reality games for teaching security concepts to first-year undergraduates. In 8th Workshop on Cyber Security Experimentation and Test ({CSET} 15).
- GRS, Government of Serbia. 2017. Strategy for the Development of Information Security in the Republic of Serbia for the period from 2017 to 2020. [ONLINE] Available at: <https://ials.sas.ac.uk/eagle-i/official-gazette-republic-serbia>. [Accessed 5 February 2020].
- Higgins, K., 2018. Best practices for recruiting & retaining women in security. Dark Reading. Information Week.
- ISACA. (2014). The growing cybersecurity skills crisis: Addressing the conflict of too many threats, too few skilled professionals. [ONLINE] Available at: [http://www.isaca.org/cyber/Documents/Cybersecurity-Report\\_pre\\_Eng\\_0414.pdf](http://www.isaca.org/cyber/Documents/Cybersecurity-Report_pre_Eng_0414.pdf) [accessed 13 April 2020]
- ITU. 2019. Cybersecurity/Trust; How ITU and the Republic of North Macedonia collaborate to strengthen cybersecurity. [ONLINE] Available at: <https://news.itu.int/how-itu-and-the-republic-of-north-macedonia-collaborate-to-strengthen-cybersecurity/>. [Accessed 17 April 2020].
- Jethwani, M.M., Memon, N., Seo, W. and Richer, A., 2017. "I Can Actually Be a Super Sleuth" Promising Practices for Engaging Adolescent Girls in Cybersecurity Education. *Journal of Educational Computing Research*, 55(1), pp.3-25.
- Kosovo Live, 2016, women are breaking the stereotype that the field of ICT belongs only men. [ONLINE] Available at: <http://kosalive.org/femrat-po-e-theyjne-stereotipin-se-fusha-e-tik-ut-u-perket-vetem-meshkujve/> [Accessed March 2020]
- LeClair, J., Shih, L. and Abraham, S., 2014, February. Women in STEM and cyber security fields. In Proceedings of the 2014 Conference for Industry and Education Collaboration (pp. 5-7).
- Maraj, A., Jakupi, G., Rogova, E. and Grajqevci, X., 2017, June. Testing of network security systems through DoS attacks. In 2017 6th Mediterranean Conference on Embedded Computing (MECO) (pp. 1-6). IEEE.
- Maraj, A., Rogova, E. and Jakupi, G., 2020. Testing of network security systems through DoS, SQL injection, reverse TCP and social engineering attacks. *International Journal of Grid and Utility Computing*, 11(1), pp.115-133.
- Maraj, A., Rogova, E., Jakupi, G. and Grajqevci, X., 2017, October. Testing techniques and analysis of SQL injection attacks. In 2017 2nd International Conference on Knowledge Engineering and Applications (ICKEA) (pp. 55-59). IEEE.
- Master, A., Cheryan, S. and Meltzoff, A.N., 2016. Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of educational psychology*, 108(3), p.424.
- Master, A., Cheryan, S., & Meltzoff, A. (2015). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology*, 108(3), 424–437. doi: 10.1037/edu0000061
- MEST. 2018. Education statistics in Kosovo. [ONLINE] Available at: <https://masht.rks-gov.net/uploads/2018/07/statistikate-arsimit-ne-kosove-2017-18.pdf>. [Accessed 18 February 2020].
- MHMR - Ministry of Human and Minority Rights. 2019. REPORT OF MONTENEGRO ON THE IMPLEMENTATION OF THE BEIJING DECLARATION AND PLATFORM FOR ACTION (BPfA) AND 2030 AGENDA FOR SUSTAINABLE DEVELOPMENT

- (2030 AGENDA). [ONLINE] Available at: [https://unece.org/fileadmin/DAM/RCM\\_Website/Montenegro.pdf](https://unece.org/fileadmin/DAM/RCM_Website/Montenegro.pdf). [Accessed 29 April 2020].
- Minović, A., Abusara, A., Begaj, E., Erceg, V., Tasevski, P., Radunović, V., Klopfer, F. and DiploFoundation, G., 2016. Cybersecurity in the Western Balkans: Policy gaps and cooperation opportunities. Research Report, 2016. Accessed January 12, 2017. <https://www.diplomacy.edu/sites/default/files/Cybersecurity%20in%20Western%20Balkans.pdf>.
- Moorman, P. and Johnson, E., 2003. Still a stranger here: Attitudes among secondary school students towards computer science. ACM SIGCSE Bulletin, 35(3), pp.193-197.
- Morgan, S., 2019. Cybersecurity talent crunch to create 3.5 million unfilled jobs globally by 2021. Cybercrime Magazine.
- OSCE. 2018. Irina Rizmal, Guide through information security in the Republic of Serbia. Belgrade, Unicom Telecom, Belgrade, IBM, Juniper, 2018, ISBN 978-86-6383-078-3. [ONLINE] Available at: <https://www.osce.org/serbia/272171>. [Accessed 1 April 2020].
- Peacock, D. and Irons, A., 2017. Gender inequality in cybersecurity: Exploring the gender gap in opportunities and progression. International Journal of Gender, Science and Technology, 9(1), pp.25-44.
- Poposka, V., 2016. THE URGE FOR COMPREHENSIVE CYBER SECURITY STRATEGIES IN THE WESTERN BALKANS. Information & Security, 34(1), pp.25-36.
- Reed, J., Zhong, Y., Terwoerds, L. and Brocaglia, J., 2017. The 2017 global information security workforce study: Women in cybersecurity. Frost & Sullivan White Paper.
- Rowland, P., Podhradsky, A. and Plucker, S., 2018, January. CybHER: A Method for Empowering, Motivating, Educating and Anchoring Girls to a Cybersecurity Career Path. In Proceedings of the 51st Hawaii International Conference on System Sciences.
- Schurr A., The Best Offense Is A Diverse Defense, Diversity Professional. 2020 [ONLINE] Available at: [https://mydigitalpublication.com/publication/frame.php?i=310137&p=&pn=&ver=html5&view=articleBrowser&article\\_id=2508811](https://mydigitalpublication.com/publication/frame.php?i=310137&p=&pn=&ver=html5&view=articleBrowser&article_id=2508811). [Accessed 4 November 2020].
- Šendelj, R. and Ognjanović, I., 2015. Cyber Security Education in Montenegro: current trends, challenges and open perspectives. In The 7th annual International Conference on Education and New Learning Technologies (EDULEARN15).
- Seymour, E., 1995. Guest comment: Why undergraduates leave the sciences. American Journal of Physics, 63(3), pp.199-202.
- STIKK, Kosovo Information and Communication Technology Association, Challenges of women in the field of ICT, 2014. [ONLINE] Available at: [http://stikk.org/fileadmin/user\\_upload/femrat\\_ne\\_teknologiji\\_-\\_sfidat\\_e\\_femrave\\_ne\\_tik.pdf](http://stikk.org/fileadmin/user_upload/femrat_ne_teknologiji_-_sfidat_e_femrave_ne_tik.pdf). [Accessed April 2020]
- USEJOURNAL. 2018. The Importance of Women in Technology, Liat Portal. [ONLINE] Available at: <https://blog.usejournal.com/the-importance-of-women-in-technology-15a653d12c>. [Accessed 12 February 2020].
- World Bank Report, ICT Perspective for Young Women in Rural Areas, October 28, 2015, [ONLINE] available at: <https://www.worldbank.org/en/news/press-release/2015/10/28/kosovo-world-bank-ict-perspective-for-young-women-in-rural-areas>. [Accessed 17 April 2020]

# IoT Security and Forensics: A Case Study

Erik David Martin<sup>1</sup> Iain Sutherland<sup>2</sup> and Joakim Kargaard<sup>3</sup>

<sup>1</sup>Sopra Steria, Stavanger, Norway

<sup>2</sup>Noroff University College, Elvegata 2A, Norway

<sup>3</sup>Noroff Education Elvegata 2A, Norway

[Erik.martin@soprasteria.com](mailto:Erik.martin@soprasteria.com)

[Iain.sutherland@noroff.no](mailto:Iain.sutherland@noroff.no)

[Kim.kargaard@noroff.no](mailto:Kim.kargaard@noroff.no)

DOI: 10.34190/EWS.21.032

**Abstract:** The expansion of the Internet of Things (IoT) has resulted in a corresponding increase in the amount of data contained in IoT devices. This data relates either to device-use or the local environment and provides digital forensic investigators with sources as diverse as smart cameras, sensors and watches with an opportunity to potentially capture or corroborate evidence. A challenge is that the data collected as part of the ecosystem operation may be distributed between several locations: cloud storage, mobile phone applications and the physical IoT device. Conducting digital forensic investigations on these devices is demanding. The hardware, mobile phone applications and overall design are, in most cases, unique to the manufacturer and model. In some scenarios, physically accessing devices can provide an opportunity to extract forensically valuable data. According to previous studies by OWASP's top 10 IoT project, many IoT devices are vulnerable. This includes the digital and physical security aspects of the device. There have been numerous studies on individual devices, many using a vulnerability to access the device. Although vulnerabilities are often quickly patched, it is noticeable that several jurisdictions are inducing legal requirements for IoT devices to be secured. This is due to growing security concerns and the number of insecure IoT devices on the market. A popular IoT device was selected and examined, highlighting the importance of applying an appropriate forensic process when physically accessing these devices. The focus was on extracting forensic data using various appropriate methods, in this case, probing the Universal Asynchronous Receiver-Transmitter (UART) ports. The test device was accessed using UART, and it was then possible to discover the device's root password. The password was hardcoded and was not uniquely generated for every device, highlighting the security concerns targeted in recent legislation. Findings throughout the research allow a digital forensic investigator to access and extract data. This could potentially increase the chance of obtaining relevant evidence during an investigation.

**Keywords:** IoT, cybersecurity, data extraction, firmware, computer forensics, OWASP

---

## 1. Introduction

The variety and volume of the Internet of Things (IoT) devices continue to increase as new markets develop and domestic devices' costs decrease. Manufacturers were fixed on market share and price while developing IoT devices, and, unfortunately, the focus on security was limited (Brumfield, 2020).

Security concerns are addressed by new legislation that has come into force in some jurisdictions, including the State of California (California State legislature, 2018), the State of Oregon (Oregon State legislature, 2018) and the Cybersecurity Improvement Act 2020 (US Congress 2020) in the USA. Also, guidance on best practice such as the NIST IoT Device Cybersecurity Guidance for the Federal Government (Fagan, Marron, Brady, *et al.*, 2020) and the NIST Foundational Cybersecurity Activities for IoT Device Manufacturers (Fagan, Megas, Scarfone, *et al.*, 2020a, 2020b).

IoT devices can be integrated into their environment and gather increasing amounts of data relating to the local environment or device used by the individual(s). Unfortunately, these devices can also offer a potential access point for those with less honourable intentions. It is possible to access sensitive data since the compromise of one device in the local IoT network can also affect other devices. (Martin, Kargaard and Sutherland, 2019). Smart cameras, sensors, watches and other IoT devices have the potential to capture or corroborate evidence (MacDermott, Baker, and Shi, 2018). Digital forensic investigators are then provided with devices that can potentially provide valuable digital evidence from a scene, possibly as a tool, a target or as a witness of an incident (Li, *et al.*, 2015).

The data collected from an IoT ecosystem operation may be distributed between several locations; the IoT device may contain the information required as part of its operation, it may be controlled via a smartphone application and may transfer information to the user's account in the manufacturer or possibility on a third-party cloud storage system (Alenezi *et al.*, 2019). Due to the drive for interoperability, one device system is likely

to communicate with an extensive ecosystem to enable control, for instance, by a Home Hub (Awasthi *et al.*, 2018).

## **2. Challenges of forensics on IoT devices**

Digital forensic investigations on IoT systems present several challenges. The hardware, mobile phone applications and device configuration may be unique to a manufacturer, a model, a firmware version, and some more complex devices with firmware updates and customisation, unique to the individual device (Atlam *et al.*, 2020). Some suggest that accessing the controlling phone application may contain relevant data regarding IoT devices (Alenezi *et al.*, 2019). However, physically accessing these devices can provide an opportunity to extract valuable data not available inside the controlling application.

Device access can possibly be provided via a login or may require more extensive work to identify a possible avenue to access the information stored in the device. There have been numerous studies on individual devices, many using vulnerabilities in the device design or firmware/software implementation (Mahbub, 2020).

According to previous studies by OWASP's (2018) top 10 IoT project, several IoT devices are vulnerable. This includes both digital and physical security aspects of the device. However, vulnerabilities may often be patched quickly by the manufacturer, improving device security but closing a potential avenue for a forensic investigator to gain access to valuable data. Physical access to the device allows the opportunity to exploit several possible routes to access data, including Universal Asynchronous Receiver-Transmitter (UART), Joint Test Action Group (JTAG) and a more invasive chip-off approach to interrogate memory.

## **3. Existing guidelines and methodologies for IoT security and forensics**

Both NIST and ISO have published several guidelines and standards relating to cybersecurity and forensics to guide how to extract information from IoT and mobile devices during an investigation, including; NIST Special Publication 800:213 (Fagan, Marron, Brady, *et al.*, 2020), the NIST Report 8259 on Foundational Cybersecurity Activities for IoT Device Manufacturers (Fagan M., Megas K., Scarfone *et al.*, 2020a), also the IoT Device Cybersecurity Capability Core Baseline (Fagan M., Megas K., Scarfone *et al.*, 2020b), ISO 30141:2018, ISO/IEC 30141:2018, ISO 27037:2012, and National Institute of Standards and Technology (2020).

The challenge with the ISO Standard is that it was published in 2012, within a fast-moving area with changes in number and types of devices. The NIST SP 800-101 publication, which provides Mobile Device Forensics guidelines, was published in 2014. (Ayers, Brothers, and Jansen, 2014) The focus is on manual extraction, logical extraction, hex dumping, chip-off and micro read classification systems. The ISO 27037:2012 Standard provides information for the identification, collection, acquisition and preservation of evidence.

OWASP's firmware security testing methodology provides nine steps to obtain and analyse embedded devices' firmware (OWASP, 2020; OWASP, 2019). This also applies to IoT devices. OWASP's methodology discusses how the firmware can be obtained by physically accessing the device using techniques such as UART and JTAG.

However, physically accessing the device to conduct firmware extraction is not always necessary. The firmware can, in some cases, be published by the manufacturing team or third parties. Other techniques, such as a Man-In-The-Middle (MITM) attack, can reveal the firmware's download URL, as there is a lack of secure firmware upgrade mechanisms. The same applies when connecting to the serial communication port, as sensitive information can be leaked (OWASP, 2020).

Moreover, the data stored locally on the device can be obtained using insecure network services, such as Telnet or File Transfer Protocol (FTP). Many IoT devices have these services exposed externally and have weak login credentials (Martin, Kargaard and Sutherland, 2019; OWASP, 2018). Therefore, these services can be used to conduct remote access and obtain internal data. However, this paper's research shows a new trend towards IoT manufacturers. They have turned off these services externally instead of exposing them on the network, indicating adherence to best practices. In practice, a forensic examiner must gain physical access to the device to conduct data extraction. Manufacturers have created various new IoT devices, used across many different industries and in many ways, making it difficult to provide one IoT definition (Megas, Piccarreta and O'Rourke, 2017). Due to the speed to market, many devices are not secured correctly. The US government departments have started using IoT devices as tools for asset tracking, monitoring and access control (Pingol, 2020). However,

due to the security issues inherent within devices, the US government has set up a new law to ensure that government departments do not fall prey to these vulnerabilities, and NIST (The National Institute of Standards and Technology) is required to put in place a new standard for IoT security, which government departments must adhere to (ibid).

This new law indicates that the standards that have already been put in place are no longer enough to ensure compliance. California Law has gone further and is compelling all manufacturers that sell connected devices in California to ensure they are secured through a unique password or the ability to change the authentication method once logging in for the first time (George, 2020).

INTERPOL has also highlighted concerns over the security of IoT devices. The 2018 INTERPOL Digital Security Challenge focused on IoT devices and a compromised webcam scenario (INTERPOL, 2018), suggesting that best practices currently recommended are not being followed. These devices are an area of concern. This paper's case example highlights a device made available for sale in 2019 to the global public. It showcases the concerns and issues with compliance to best practice.

## **4. IoT device example**

### **4.1 Device selection**

In the light of the new legislation and recommended best practice, an IoT device was selected for analysis. TP-Link is a popular and well-known vendor. They offer a series of products, including IoT and routers, and in late 2019 a new camera called Tapo C200 was released. The camera was sold throughout 2019 and 2020 and is popular on marketplaces such as Amazon (2021). The camera was also chosen for this case study as it was highlighted as having security issues that were indicated to the manufacturer and subsequently fixed (Which, 2020). A TP-Link C200 IoT camera, firmware build 1.0.14, was purchased in September 2020 and later updated to firmware build 1.0.17, as of February 2021. The camera offers Wi-Fi connectivity and a mobile phone application. The application has a series of features. This includes video recording and streaming, motion detection and two-way audio communication. The camera has a small internal storage for its firmware and an external micro-SD card slot. The SD card can be used to store videos and photographs. Also, the vendor provides the source code online for its consumers, which is under the GNU General Public License (TP-Link, 2021).

### **4.2 Camera testing methodology**

OWASP's (2020) firmware security testing methodology is followed when conducting forensic analysis. The camera PCB has four pins, VCC, TX, RX and GND. These pins are used for Universal Asynchronous Receiver-Transmitter (UART) communication which manufacturers can use for debugging purposes. UART also allows a forensic investigator to connect physically to the device and gain remote access. The device was accessed using UART. From there on, it was possible to discover the device's internal storage due to hardcoded credentials. The internal storage includes data, such as the device's Wi-Fi password, timestamps and public IP address. The data can be exfiltrated using the binaries included on the device itself, such as Netcat.



**Figure 1:** Probing kit connected to UART



### 4.3 Connecting to the device

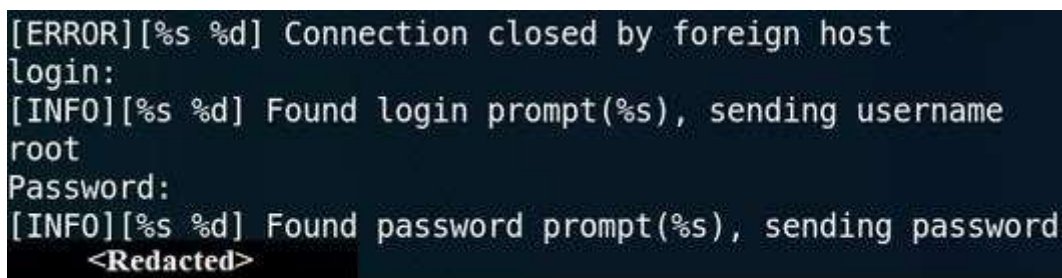
UART communication is achieved using a probing kit connected to an FTDI 3.3-volt breakout board. This allows serial communication between the camera and the board. The breakout board is then connected directly to the investigator's computer using a USB connection. Once the board is connected to the computer, a new device called ttyUSB0 is available. The application screen can be used to establish communication between the computer and the camera on the Linux operating systems (Die, 2003). An application such as PuTTY can also be used for the same purpose.

Furthermore, the correct baud rate must be set between the camera and the computer. Baud rate is the measurement of the data transfer speed defined in bits per second (bps) (Campbell, 2016). The baud rate is mostly unique depending on the device, as there are several standard rates. The camera's baud rate is 57600 bps, which was discovered after connecting with a series of standard rates. If the baud rate is incorrect, the camera's output cannot be interpreted.

Once connected to the camera, the investigator can interact directly with the camera's shell. The device is protected with a username and password. However, previous research shows that many IoT devices have weak and hardcoded credentials (OWASP, 2018). A series of default passwords for the root account did not work for this specific device. However, debug messages were suddenly displayed in the console when connected using UART. The same applied when interacting with the camera's mobile application. This is an indication that most camera interaction is displayed when using the camera's shell. This gives the investigator an opportunity to both view and access the firmware location when upgrading the device, as mentioned in the OWASP's (2020) firmware security testing methodology.

When upgrading the camera's firmware, its download location is revealed as a debug message. An investigator can download the firmware and determine the hardcoded root password if there is one. However, if the download URL is not available in the console, it is possible to sniff the network traffic and reveal the location using tools such as Wireshark or Tcpdump. The same applies to conducting a man-in-the-middle attack (OWASP, 2020). After gaining access to the firmware, binwalk was used to extract the file system. The file system could now be analysed locally on the investigator's computer. The /etc/passwd file contained the root password hashed with md5crypt. This can potentially be cracked with tools such as Hashcat or John the Ripper.

### 4.4 Gaining access and data extraction



```
[ERROR][%s %d] Connection closed by foreign host
login:
[INFO][%s %d] Found login prompt(%s), sending username
root
Password:
[INFO][%s %d] Found password prompt(%s), sending password
<Redacted>
```

Figure 2: The credentials discovered inside the /usr/bin/logrecord file<sup>1</sup>.

Moreover, a binary called /usr/bin/logrecord was discovered. This is used for logging the camera's activity. Thus, the binary contains the root password in cleartext. Using the application 'strings' against the binary reveals the hardcoded password. The investigator can now successfully log into the system and create a telnet backdoor, which allows a remote shell across the network rather than physically connecting to the device with UART. This could also be misused from an attacker's point of view. If the shell did not have root permissions, the user could privilege escalate using, for example, kernel exploits, as the kernel is outdated.

<sup>1</sup>TP-Link was informed of the vulnerability; however, this issue still exists in the latest version of the firmware. Therefore, the password was redacted in figure 2. Law enforcement should contact the lead author from an official email account if this information is needed as part of an ongoing investigation.

```
root@SLP:~# id
uid=0(root) gid=0(root) groups=0(root)
root@SLP:~# uname -a
Linux SLP 3.10.27 #1 PREEMPT Mon Jul 20 10:42:22 CST 2020 rlx GNU/Linux
root@SLP:~# /usr/sbin/telnetd -b 192.168.0.113
```

**Figure 3:** Root shell inside the camera and opening a telnet backdoor

The file `luci-sessions` contained information regarding the private IP address used to connect to the camera through the mobile application. This would link the user's mobile phone to the IoT device. Similar timestamps can be located in the `diagnose_log` file. Even though the SD card contains videos and images, the camera's internal storage contains information such as the wireless network's SSID, Wi-Fi password, public IP address and logged user activity. The Wi-Fi password is stored in cleartext in the `wpa_supplicant.conf` file.

Suppose the router's Wi-Fi password has been set manually. In that case, the cleartext Wi-Fi credentials on the device could potentially lead to accessing other services and devices by the same user, as password reuse is common (Mirkovic *et al.*, 2016).

An investigator can now extract data stored locally on the device. Many IoT devices' firmware includes a series of binaries by using BusyBox. BusyBox is an all-in-one package containing many of the most common UNIX binaries (BusyBox, 2008). The C200 camera includes BusyBox. Therefore, an investigator could transfer internal files across the network using netcat, for example.

## 5. Discussion

The internal storage contains valuable information regarding both the user activity and wireless network settings. Connecting to the camera using UART will output any debug information. While connected, an investigator can upgrade the camera's firmware using the device's mobile application. This will reveal the firmware download URL.

This methodology can be used on other IoT devices as well, as stated by OWASP (2020). Obtaining the firmware location allows an investigator to download and examine the content, potentially revealing the hardcoded credentials. In practice, the investigator could get valuable evidence on the device internally, and not only from connected external storage devices such as SD cards. Several files located internally were not found on the external storage device. The internal data can be extracted across the network by taking advantage of the BusyBox utilities.

TP-Link was informed regarding the hardcoded credentials and the `logrecord` binary. The vendor released a new firmware upgrade (the current version at the time of writing - build 1.0.17 - latest). However, the new version still had the hardcoded credentials, as the password was not uniquely generated or permanently removed according to best practice. This highlights the issue that manufacturers are not adhering to guidelines, standards or laws and do not always realise the severity of the problem, despite being informed of the issue.

## 6. Summary

The Internet of Things (IoT) continues to expand. As a result, there is a corresponding increase in the amount of data gathered by these devices relating to their use or the local environment. This provides an ideal opportunity for digital forensic investigators since smart cameras, sensors and watches can capture or corroborate evidence.

A challenge is that the data collected as part of the ecosystem operation could be distributed between several locations: cloud storage, mobile phone applications and the physical IoT device. Conducting digital forensic investigations on these devices is demanding.

The hardware, mobile phone applications and overall design are, in most cases, unique to a manufacturer, a model and in some cases, to the individual device. In some cases, physically accessing these devices can provide an opportunity to extract unique data, which is otherwise inaccessible on the mobile application or the external storage device.

There have been numerous studies on individual devices using different techniques to access the device due to a vulnerability. According to previous studies by OWASP's top 10 IoT project, many of these devices are considered vulnerable. This includes both the digital and physical security aspects of devices. However, vulnerabilities are often quickly patched.

Insecure devices are often beneficial for forensic examiners, as the internal data can be accessed easily. This has been demonstrated in this paper. However, this is also a security issue from a consumer point of view. This paper describes a possible process for accessing these devices to extract data using various methods, including Universal Asynchronous Receiver-Transmitter (UART).

This paper focuses on the security issues and the need for a digital forensics' methodology regarding IoT devices. TP-Link's C200 IoT camera is used as an example of how to extract data through physical access. The device was accessed using UART, which resulted in firmware extraction. From there on, it was possible to discover the device's hardcoded root password.

## **7. Conclusion**

The same methodology for accessing firmware can be implemented on other IoT devices by following OWASP's firmware security testing methodology. Upgrading the device's firmware while monitoring the device's output can reveal the firmware download location. Examining the firmware allows an investigator to discover hardcoded credentials and access to the device's internal storage.

The internal data extracted from TP-Link's C200 camera includes the SSID, Wi-Fi password, public IP address and user interaction with timestamps. This indicates that similar devices might include the same amount of data. Suppose the Wi-Fi password on the router has been manually set. In that case, the cleartext password discovered on the camera could give access to other services and devices owned by the same user, as password reuse is still common.

In conclusion, IoT devices could contain valuable evidence both in the internal and external storage device. Having physical access to these devices could lead to obtaining otherwise inaccessible evidence. The OWASP top 10 IoT project confirms that hardcoded credentials are still common. This has also been documented in this paper, as the root password is not uniquely generated across every TP-Link C200 camera. Therefore, computer forensics practitioners should also reserve time to access the device's internal storage and external data sources.

## **References**

- Alenezi, A., Atlam, H., Alsagri, R., Alassafi, M.O., Wills, G., (2019) *IoT Forensics: A State-of-the-Art Review, Challenges and Future Directions*. The 4th International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS 2019). 10.5220/0007905401060115.
- Amazon. (2021) TP-Link Tapo Smart Cam Pan Tilt Home Wi-Fi Camera. <https://www.amazon.com/TP-Link-Tapo-Wireless-Security-C200/dp/B0829KDY9X>.
- Atlam, H., Alenezi, A., Alassafi, M.O., Alshdadi, A. A., Wills, G., (2020) Security, cybercrime and digital forensics for IoT. In book: *Principles of Internet of Things (IoT) Ecosystem: Insight Paradigm* (pp.551-577) DOI: 10.1007/978-3-030-33596-0\_22.
- Awasthi, A., Read, H.O.L, Xynos, K., Sutherland, I., (2018) *Forensic Analysis of the Almond+ Smart Home Hub Environment: First Impressions, Digital Forensics Research Workshop (DFRWS) -USA Rhode Island, July 15-18, 2018*.
- Ayers R., Brothers S., Jansen W., (2014) *Guidelines on Mobile Device Forensics*. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-101, Rev. 1. <http://dx.doi.org/10.6028/NIST.SP.800-101r1>
- Brumfield, C. (2020) New US IoT law aims to improve edge device security. [online] CSO Online. Available at: <https://www.csoonline.com/article/3597956/new-us-iot-law-aims-to-improve-edge-device-security.html>
- BusyBox. (2008) BusyBox: The Swiss Army Knife of Embedded Linux. <https://www.busybox.net/about.html>
- California State legislature (2018) SB327 <https://www.natlawreview.com/article/iot-manufacturers-what-you-need-to-know-about-california-s-iot-law>
- Campbell, S. (2016) Basics of UART communication. Circuit Basics. <https://www.circuitbasics.com/basics-uart-communication/>
- Die. (2003) screen (1) - Linux man page. <https://linux.die.net/man/1/screen>
- Fagan M., Marron J., Brady K., Cuthill B., Megas K., Herold R. (2020a) IoT Device Cybersecurity Guidance for the Federal Government: Establishing IoT Device Cybersecurity Requirements. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-213, Draft. <https://doi.org/10.6028/NIST.SP.800-213-draft>

- Fagan M., Megas K., Scarfone K., Smith M. (2020a) Foundational Cybersecurity Activities for IoT Device Manufacturers (National Institute of Standards and Technology, Gaithersburg, MD), NIST Intragency or Internal Report 8259. <https://doi.org/10.6028/NIST.IR.8259>
- Fagan M., Megas K., Scarfone K., Smith M. (2020b) IoT Device Cybersecurity Capability Core Baseline (National Institute of Standards and Technology, Gaithersburg, MD), NIST Intragency or Internal Report 8259A. <https://csrc.nist.gov/publications/detail/nistir/8259a/final>
- George, D., (2020) IoT Manufacturers – What You Need to Know About California's IoT law. [online] The National Law Review. Available at: <https://www.natlawreview.com/article/iot-manufacturers-what-you-need-to-know-about-california-s-iot-law>
- INTERPOL (2018) 'Internet of Things' cyber risks tackled during INTERPOL Digital Security Challenge. <https://www.interpol.int/News-and-Events/News/2018/Internet-of-Things-cyber-risks-tackled-during-INTERPOL-Digital-Security-Challenge>
- ISO/IEC 27030 — Information technology — Security techniques — Guidelines for security and privacy in the Internet of Things (IoT) [DRAFT] <https://www.iso27001security.com/html/27030.html>
- ISO 27037:2012 Information technology — Security techniques — Guidelines for identification, collection, acquisition and preservation of digital evidence <https://www.iso.org/standard/44381.html>
- ISO/IEC 30141:2018 Internet of Things (IoT) – Reference Architecture, <https://www.iso.org/obp/ui/#iso:std:iso-iec:30141:ed-1:v1:en>
- Li, S., Choo, K.K.R., Sun, Q., Buchanan, W.J., Cao, J., (2015) IoT forensics: Amazon echo as a use case. IEEE Internet of Things Journal, 6(4), pp.6487-6497. <https://uwe-repository.worktribe.com/preview/850286/iotfR2.pdf>
- MacDermott, A., Baker, T., Shi, Q. (2018) *IoT forensics: Challenges for the IOA era*. In 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS) (pp. 1-5). IEEE.
- Mahbub, M. (2020) Progressive Research on IoT Security: an exhaustive analysis from the perspective of protocols, vulnerabilities, and pre-emptive architectonics. In: *Journal of Network and Computer Applications, Volume 168:15*.
- Martin, E. D., Kargaard, J. and Sutherland, I. (2019) Raspberry Pi Malware: An Analysis of Cyberattacks Towards IoT Devices. 10th International Conference on Dependable Systems, Services and Technologies (DESSERT).
- Megas, K., Piccarreta, B. and O'Rourke, D., (2017) *Internet of Things (IoT) Cybersecurity Colloquium*. Arlington: National Institute of Standards and Technologies.
- Mirkovic, J. Hanamsagar, A. Kanich., C., Woo, S.S. (2016) *How Users Choose and Reuse Passwords*, University of Southern California, Information Science Institute Technical Report 75 (ISI-TR-715) November 2016.
- National Institute of Standards and Technology (2020) *IoT Security Related Initiatives at NIST* Available at: <https://www.nist.gov/itl/applied-cybersecurity/nist-initiatives-iot>
- OWASP. (2020) OWASP Firmware Security Testing Methodology. <https://scriptingxss.gitbook.io/firmware-security-testing-methodology/>
- OWASP. (2019) OWASP Internet of Things Project. [https://wiki.owasp.org/index.php/OWASP\\_Internet\\_of\\_Things\\_Project#tab=Firmware\\_Security\\_Testing\\_Methodology](https://wiki.owasp.org/index.php/OWASP_Internet_of_Things_Project#tab=Firmware_Security_Testing_Methodology)
- OWASP. (2018) Top 10 IoT Project. <https://owasp.org/www-pdf-archive/OWASP-IoT-Top-10-2018-final.pdf>
- Oregon State legislature (2018) Oregon House Bill 2395 amending ORS 646.607. <https://www.natlawreview.com/article/oregon-s-new-iot-law>
- Pingol, E., (2020) US IoT Improvement Act Becomes Law. [online] IoT Security. Available at: <https://www.trendmicro.com/us/iot-security/news/6613>
- TP-Link (2021) GPL Code Center. <https://www.tp-link.com/us/support/gpl-code/>
- US Congress (2020) H.R.1668 - IoT Cybersecurity Improvement Act of 2020 <https://www.congress.gov/bill/116th-congress/house-bill/1668/actions>
- Which (2020) TP-Link camera security flaw discovered in Which tests as IoT law moves closer <https://www.which.co.uk/news/2020/07/tp-link-camera-security-flaw-discovered-in-which-tests-as-iot-law-moves-closer/>

# Cybersecurity and local Government: Imperative, Challenges and Priorities

Mmalerato Masombuka<sup>1</sup>, Marthie Grobler<sup>2</sup> and Petrus Duvenage<sup>3</sup>

<sup>1</sup>University of Stellenbosch, South Africa

<sup>2</sup>CSIRO's Data61, Melbourne, Australia

<sup>3</sup>Academy of Computer Science and Software Engineering, University of Johannesburg, South Africa

[mmalerato.mc@gmail.com](mailto:mmalerato.mc@gmail.com)

[marthie.grobler@data61.csiro.au](mailto:marthie.grobler@data61.csiro.au)

[duvenage@live.co.za](mailto:duvenage@live.co.za)

DOI: 10.34190/EWS.21.501

**Abstract:** The South African government's pursuit of widespread internet access, its increasing use of and reliance on digital services as well as the emergence of new technologies have given rise to new threats and risks of cyberattacks. An effective cybersecurity approach in countering these threats requires a coherent effort involving all spheres of government. However, given the government's three-tiered structure (national, provincial, and local government), there seems to be disproportionality in the cybersecurity approach. Without distracting from the importance of cybersecurity at a national level, it is imperative for cybersecurity at provincial and local government to be prioritised, resourced and fit for purpose. While there are commonalities between these spheres, the contexts within which provincial and local government functions differ from the national. Thus, the one size fits all cybersecurity approach being employed by the national government is neither sufficiently inclusive nor fully downward scalable. The continuous evolution of cyberspace and the associated threats call for a concurrent and continuous adaptation in the approaches employed to build resilient cybersecurity on all levels of government. If not, local government, in particular, will continue to be an attractive target and a weak chink in the government's cybersecurity armour. Therefore, this paper aims to contribute to the discourse on cybersecurity at the local government level. In our focus on the insecurity that local governments face in the cyber domain, we use South Africa as an illustrative example. The paper discusses the imperative of raising the cybersecurity bar at local government level and examines the challenges in this regard. We then proceed with proposing key priorities that local government could implement to enhance cyber security - as part of which we explore the application of the Australian Signals Directorate's Cyber Security Centre's Essential Eight Maturity Model and the NIST Cyber Security Framework (CSF) to local government.

**Keywords:** cybersecurity, cyberthreats, cyber resilience, local government, National Cybersecurity Policy Framework (NCPF)

---

## 1. Introduction

The security of cyberspace remains a growing concern for governments around the world. This concern is largely driven by the pursuit of universal access to the internet, the growing reliance on e-government services, and the emergence of new technologies and technological evolutions. Our work focuses on the insecurity that local governments face in the cyber domain, and uses South Africa as an illustrative example to showcase the imperatives, challenges and opportunities faced by local governments.

There is limited research focused on cybersecurity at the local government level in Africa, in particular South Africa. Thus, the paper not only aims to elevate the prominence of the discourse on cybersecurity at the grassroots level but aims to lay the foundation for future work in this developing domain. The first part of the paper provides context on cybersecurity at the local government level. Local governments continue to be a soft target for cyber adversaries due to the amount of constituent data that they possess. Consequently, the third section aims to demonstrate the significance of improving cybersecurity at local level. The fourth part highlights challenges to bolstering cybersecurity. Lastly, in seeking to develop cyber resilience at local government, we look at priorities for cyber resilience and cybersecurity framework.

## 2. Local government and cybersecurity: Concept and context

This section contextualises cybersecurity at local government level with specific reference to South Africa. The section also provides a general definition of local government, describes the position, role and functions of local government in South Africa and overviews the responsibility and status of cybersecurity at the local government level.

## **2.1 The general definition of local government**

The term *local government* has been widely and variedly defined by scholars and political systems. For this paper, we adopt the definition that local government denotes the “government of a specific local area constituting a subdivision of a major political unit such as a nation or state” (Merriam-Webster Online). Typically, a local government has the responsibility of administering and providing public services and facilities in a particular area. In attempting a comprehensive definition of local government, Ndreu (2016:7) describes local government as a combination of several elements that include the existence of a local authority regulated by law, provision of public services for residents of the area within its jurisdiction, the local population and a defined territory, with an autonomy limited by the central government and a separation of local and non-local functions.

A local authority's degree of autonomy, the powers it exerts and the exact functions it fulfils are highly dependent on the political dispensations within which it functions. The next subsection contextualises 'local government' within the South African political dispensation.

## **2.2 Local government in South Africa**

The Constitution of the Republic of South Africa (1996) establishes three spheres of government – national, provincial and local level. The local sphere of government consists of municipalities, which has “the right to govern, on its own initiative, the local government affairs of its own community, subject to national and provincial legislation” (Constitution of the Republic of South Africa 1996). South Africa's 278 local authorities not only generate their income (through inter alia property taxes, levies, duties and service charges, etc.) but also receive financial support from the national government. These local authorities vary considerably in size and comprise eight metropolitans, 44 district and 226 local municipalities (Mokoena, 2019).

The overall responsibilities of local governments include the delivery of water and electricity, municipal health services, fire-fighting services, municipal transport infrastructure and traffic control, municipal security services, sewage and sanitation, recreational areas etc. Within South Africa, as is the case in numerous other countries, local government is essential for effective governance and service delivery. Effective governance and service delivery on a local government level are increasingly data-driven and cyber dependent. Like other levels of the South African government, local authorities are increasingly adopting technology to incorporate additional functionalities, improve efficiency, increase accessibility and communication (Nyirenda-Jere and Biru, 2015). Technological advances are transforming the operations of local governments worldwide. However, as local governments include more technology within their day-to-day functionalities, using Internet-connected systems and offering more municipal services online, they not only increase their vulnerability to cyberattacks but also expose their constituents to potential cyber dangers. Moreover, the rapid hyperconnectivity, digitisation and interdependence have also introduced a level of unprecedented cyberthreats and vulnerability to a new generation of threats to the local government's information systems (Pandey, Golden, Peasley and Kelkar, 2019: 2).

## **2.3 Cybersecurity responsibility and posture**

Despite its critical importance, academic research concerning the responsibility for, and status (posture) of cybersecurity at local government in South Africa is very limited. The South African government acknowledges the significance of developing a cyber-resilient community, and therefore a National Cybersecurity Policy Framework (NCPF) was developed in 2012 to address the cybersecurity challenges faced (Gcaza, and Von Solms, 2017). This policy framework is significant as South Africa's first attempt at creating a cybersecurity-focused policy framework, yet it is a high-level policy document and does not specifically address cybersecurity at the local government level. Also, the framework is yet to achieve some of its primary objectives such as processes and fully-capacitated institutional structures and initiatives. In this case, deficient progress at the national level inevitably exacerbates the cybersecurity challenge for local government.

Cybersecurity at local government level has not been elevated as a top priority in South Africa. Eisenstein (2020: Online) asserts that, to date, no one has established national cybersecurity standards for South African local governments and in practice the functional responsibility for cybersecurity lies with each of the respective local authorities. Applicable legislation which governs the governance and admission of local authorities in general (such as the Municipal Systems Act, 2000) by extension also applies to cybersecurity aspects, but cannot be regarded as explicit cybersecurity standards called for by Eisenstein (2020: Online). Furthermore, Mabaso (2018:

90) rightly states there is a lack of strategic understanding of the municipality's cybersecurity risk profile, including the capacity of the workforce, and current threats.

It is difficult to credibly quantify and qualify the cybersecurity posture at the local government level from publicly available sources. However, and at least in as far as could be inferred from available material, the cybersecurity posture at local government is undoubtedly a cause for major concern. The South African Auditor General (AG) annually audits organs of government which include the 278 local authorities. While the AG's reports on local authorities do not measure cybersecurity specifically, the AG considers "Information Technology - Governance" (Auditor General of South Africa, 2020). The 2018/2019 AG report shows that less than 30% of local governments sufficiently comply with IT governance standards (Auditor-General South Africa, 2020). Inferring from the outcomes of the AG's report, the cybersecurity posture at local government is below standard and deficient.

With a view on establishing a premise for the rest of the paper, this section conceptualised and contextualised cybersecurity at local government level with specific reference to South Africa. In the next section, we unpack in more detail the imperative of raising the cybersecurity bar at the local government level. This discussion is done in a manner which is not only pertinent to South Africa but hopefully has wider application to local government in other countries.

### **3. Imperative of raising the cybersecurity bar (at local government level)**

Local government is not only a hub for sensitive data but it also manages and operates critical information infrastructure. Consequently, the information it possesses continues to be targeted by cyber adversaries. Moreover, the local government is often less secured in contrast to the national government. The combination of the following interrelated and overlapping factors heighten cyber risks and makes it imperative to raise the cybersecurity bar at local government level.

#### **3.1 Local government as a data hub**

Internationally local government networks will continue to be attractive targets for cybercriminals and particularly susceptible to cyberattacks because of the vast amounts of sensitive data they possess and maintain (Thompson, 2019: online). South African local government, for instance, possess data that includes citizen's data, tax data, health data, education data, voter records, driver's license details, municipal bills and other data of critical importance. Essentially, the local government acts as a collector, manager, and owner of a wealth of state data. Municipalities operate and own information infrastructures that support national critical infrastructures such as national airports, harbours, law enforcement surveillance systems, traffic cameras and Supervisory Control and Data Acquisition (SCADA) systems etc. (Mabaso, 2018: 84).

With the increasing usage of e-government and the slow shift towards the smart city, local governments will continue to accumulate great amounts of public data, some of which will be published on their intranet, social media pages or websites. It is imperative for the government to remain transparent and make public records and information easily accessible to citizens. At the same time, this transparency has made it easier for cybercriminals to exploit public systems that contain sensitive information. With the increased availability of public open data, the security of citizens' information and privacy has become a critical issue for governmental data publishing, hence local government must make cybersecurity a priority.

#### **3.2 Vulnerability of local government**

Relative to the national government and well-resourced corporates, local governments are more vulnerable and cybercriminals are capitalising on those vulnerabilities to further their personal, political, or criminal agenda. In some instances, cyber attackers exploit the smaller and more vulnerable local government to gain a foothold in the government's larger systems and subsequently attack their original target which is often a national department. Thus Osborn and Simpson (2015) argue that enhanced cyber resilience within local government will not only ensure security for small-scale cybersecurity users but it will also ensure security across all spheres of government as some of their information systems are interconnected.

With the expansion of digital technologies local governments are becoming smarter, allowing interconnection between systems, people and devices to improve infrastructure, efficiency and convenience for residents (Thompson, 2019: online). However, when adopting smart technologies, where various municipal works are

connected to a computer network or the Internet, municipalities often fail to ensure the technology is secure before implementation (Thompson, 2019: online). Concurrent with the global health crisis, 2020 has also ushered in a new age of cyberattacks aimed specifically at municipal governments. The number of local government employees who have been working from home and others who have transitioned to telecommuting as a result of the COVID-19 pandemic has introduced new network vulnerabilities and also expanded the cyberattack vector even further (Nabe, 2020).

### **3.3 Local government as softer targets**

Mcanyana et al. (2020: 6) suggest that cyber actors regularly focus on governments that are perceived as having lower defensive barriers, i.e. cyber insecurity at local government level. Cyberattacks against municipalities are increasingly common and becoming more sophisticated and severe. According to Smith (2019: online), one of the biggest trends in 2019 in South Africa were attempted cyberattacks on industrial controls systems, including dam control facilities, water and electricity facilities as well as nuclear facilities. Also in 2019, the City of Johannesburg was the victim of a ransomware attack in which a group that called itself “Shadow Kill Hackers” claimed to have control of the city’s servers and that they have dozens of back doors inside their systems. The group threatened to release the city’s stolen sensitive data on the internet if they failed to pay the demanded ransom. The Johannesburg City Power suffered a cyber-attack in 2019 and citizens were unable to access e-services including the purchase of electricity (Mcanyana, et al, 2020: 6). The malware encrypted City Power’s internal network, Web applications and official website, leaving customers without power. In 2017, it was reported that identity numbers and other financial information of approximately 60 million South African citizens (both living and deceased) were leaked on the internet while in 2016, almost 9 million South Africans fell victim to some form of cybercrime, including phishing attacks (BizNews, 2017). These cyberattacks were described as South Africa’s largest attacks ever at that time and highlighted the urgency for a broader view of cybersecurity, and how the South African Government should ensure cyber resilience in its offering of public e-services. Moreover, cyber attackers are developing more sophisticated and Artificial intelligence-based attacks that increase the speed, scale, complexity, and frequency of their attacks (Masombuka, Grobler and Watson, 2018).

### **3.4 The impact of cyber insecurity**

The impact of cyber insecurity includes significant financial loss, litigation, reputational damage and negatively impacted share prices, suspended operational activities and an eroded competitive advantage (Institute of Directors of South Africa, 2018: 3). Hubbard (2019: online) states that cybercrime is still massively underreported in South Africa, as a result, there is no reliable way of measuring the extent of financial loss or data loss or reputational risk due to cybercrime. The consequence of cyberattack at local government level could extend beyond just data loss and financial impact, instead, it could disrupt critical city services and critical infrastructure across multiple domains. The failure of the local and provincial government to ensure secure, resilient and trustworthy cyberspace does not only undermine the confidence of the information society but it also exposes the government to multiple risks and threats (Masombuka, 2018: 13).

This section demonstrated the imperative of enhancing cyber resilience at local government, provided a brief overview of the cyberthreat landscape and the implication of cyberattacks. Building on this, the next section will focus on the challenges of achieving enhanced cybersecurity at the local government level.

## **4. Challenges to bolstering cybersecurity at local government**

Some challenges to bolstering cybersecurity at the local government level are relevant to developed and developing countries alike, while other challenges are dependent on the context of the particular country. As was argued in Section 2, cybersecurity is inseparably linked with the broader management and governance of local authorities. Especially within South Africa, the management and governance of local authorities is acutely affected by socio-political, economic and other dynamics within the country. With the exception of the socio-political flux, realignment and fiscal constraints, this section highlights 10 possible barriers to achieving enhanced cyber resilience at local government. Although this list is not exhaustive, we will focus specifically on these barriers that hinder local governments in achieving enhanced cybersecurity.

Challenges to achieving enhanced cyber resilience at local government include:



#### **4.1 Insufficient funding**

Insufficient funding will increasingly prevent local governments from managing or providing the required cybersecurity protection (Norris, et al, 2020: 7). The continued under-investment on secured Information Technology systems and security initiatives will continue to leave local governments vulnerable. Microsoft (2017: 27) argues that it can be difficult to secure funding at the local government for cybersecurity unless something goes wrong and there is a financial or legal implication. Thompson (2019: online) argues that due to funding challenges, local governments use outdated technology and do not have dedicated IT staff to implement organizational safeguards to protect against the ever-increasing risk of a cyberattack.

#### **4.2 Lack of support from top officials**

Local governmental systems and databases are often the most targeted entities for cybercrime, presenting as easy targets through inadequate management and support from top officials (Toulu, 2018). This lack of support from senior management has a major impact on cybersecurity initiatives thus for the local government to achieve cybersecurity reliance, support from management is required. Mabaso (2018: 88) states that support and commitment from senior management is key and the lack thereof thus constitutes a serious barrier to achieving cyber resilience.

#### **4.3 Talent shortage**

Deloitte (2020: 1) states that there is a chronic shortage of cybersecurity talent as new technologies and evolving threats increase the level of cyber risk. To effectively deal with this shortage, local governments need to establish new recruitment methods and employ an innovative (cyber) talent framework that would inspire new and modern ways to tackle this challenge. The cybersecurity skills shortage is due to a combination of reasons, including a lack of cyber-related courses at tertiary institutions, budget constraints for formal training at local government level, rapidly changing advancements in technology and the fast-changing threat landscape.

#### **4.4 Inability to pay competitive salaries**

The inability of local governments to pay competitive salaries for cybersecurity personnel constitutes a severe barrier. Some information infrastructures require control and monitoring 24/7, particularly those that support critical infrastructures. Microsoft (2017: 6) states that it is difficult for local governments to compete with private companies in terms of cybersecurity personnel salaries and that budgetary constraints further exacerbate this barrier.

#### **4.5 Lack of capacity**

Another factor compounding the cyber risk is the identification and retaining cybersecurity personnel. Sutherland (2017: 97) states that there is an acute shortage of skilled ICT workers in South Africa, with information security a leading issue for employers. Thus, local governments need to develop a new framework that will aid in identifying talent, address capacity issues and develop retention plans.

#### **4.6 Lack of cybersecurity awareness campaigns**

The biggest cybersecurity challenge that local governments are faced with is the lack of cyber awareness and education amongst their employees, elected leaders and citizens. For instance, users (who mostly by mistake and without any malicious intent) would fall victim of a phishing attack. At a more general level, the public and all local government employees need to be taught about dangerous and unsafe behaviours on the Internet as part of a cyber-awareness drive (Sutherland, 2017: 99). To reduce cyber risks, local government leaders need to enforce cybersecurity culture, develop cyber awareness campaigns and training.

#### **4.7 No end-user accountability**

Developing a strong cybersecurity capability requires a message of accountability to users. Many workers in particular at local government fail to understand the extensive impacts of a cyberattack. Thus, Coleman (2019) states that one of the best ways of ensuring user accountability is to develop a robust model for cybersecurity accountability. The objective of this is to ensure that relevant seniors have an oversight, authority and resources to make decisions on cybersecurity that will protect the city from potential cyber threats.

#### **4.8 Lack of cybersecurity policies and practices**

Some local governments in South Africa do not have defined guidelines, practices and procedures for cybersecurity (Sutherland, 2017: 89). Such a lack of understanding of cybersecurity policies, as well as poorly enforced cybersecurity policies, will hinder cybersecurity resilience. Thus, leadership in local governments should play an important role in the cybersecurity ecosystem, not only in terms of establishing cybersecurity policies and procedures but also in terms of ensuring implementation and oversight of those programs (Deloitte, 2020: 17). Cyber adversaries often take advantage of lacking existing protocol, legal framework, and gaps of insecurity in cyber activity (Toulu, 2018).

#### **4.9 Lack of cyberculture**

Norris et al. (2020: 2) argue that top officials must play key roles in creating and maintaining cybersecurity culture throughout their government departments. This could also be achieved through the appointment and hiring of well-trained cyber experts and/or Chief Information Officers (CIOs) in various local government offices, something that the South African has a severe shortage of. Creating and enforcing a culture of cybersecurity awareness at all levels of local government is necessary to combat the evolving threat landscape (Thompson, 2019: online).

#### **4.10 The silo approach to dealing with cybersecurity-related challenges**

The silo approach to cybersecurity measures in local governments could pose a risk due to the interconnectedness of their systems. The silo approach is a barrier to achieving enhanced cyber resilience because the complexities of evolving cyber threats have crossed barriers (of ideology, politics and space), thus demanding a constructive and collaborative effort. Therefore, understanding the nature of cyberspace, its interconnectedness and interdependence will help local governments better manage cyber risks.

Every local government is unique and a fit one, fit all cyber resilience approach will be difficult to obtain. The following section highlights priorities that could be implemented by the local government to effectively respond to cyberattacks and enhance cybersecurity.

### **5. The way forward: Priorities, model and framework**

We propose some priorities that could be implemented by local governments to enhance cybersecurity. In defining government level priorities, an overarching cybersecurity framework at local government level can be adopted to suit the contexts of the local government.

In seeking to develop cyber resilience, the local government has to adequately manage cyber risks, prevent cyber incidents and minimise its impact daily (Nicholas and Pinter, 2017). The paper adopted the following priorities (Hadjizenonos, 2019: online) for actioning cybersecurity by local governments in South Africa:

#### **5.1 Prioritise cybersecurity**

The prioritisation of security measures will help local governments to strengthen cyber resilience. The pursuit of digital transformation by local government should also entail the changing of legacy systems that are obsolete and need to adapt to the new threat environments. Threats are evolving and becoming increasingly complex thus rendering legacy systems obsolete and vulnerable. In an effort to manage cyber risks, local government should use newer technologies. Local governments are also encouraged to incorporate Cyber Counterintelligence (Deloitte 2020b; Duvenage, Jaquire and von Solms, 2020) in their cybersecurity focus.

#### **5.2 Understanding the entire environment**

To combat cyberthreats, the local government must look at cyber threats across their organisation - not just as a technology issue. For example, the ability to combat a ransomware attack means identifying data that is essential to the local government and prioritising the funding to ensure that the data is protected. That goes beyond what IT and the CIO can do. It means the recognition and partnership across government to combat the threat. Cybersecurity is no longer a problem to be "solved" but an ongoing effort. The adoption of a comprehensive cybersecurity framework, such as the Cyber Security Framework (CSF) developed by the National Institute of Standards and Technology (NIST), is aimed at enhancing the security and resilience of critical infrastructure (NIST, 2018:1). A local government that lacks formal security programs can leverage the NIST

framework as a roadmap to identify security needs, prioritise investments, establish the right level of security and ultimately establish the necessary steps to address cybersecurity risks.

### **5.3 Regulations, standards and security policy**

The purpose of cybersecurity regulation is to guide governments in protecting their systems and relevant information. Such documentation allows local governments to establish frameworks for maintaining orders, handling technologies, and managing risks. This would also help employees understand what information needs to be protected, how such information should be stored and how and who should have access to it. A comprehensive example of this is the Australian Signals Directorate's Cyber Security Centre's (ACSC) Essential Eight Maturity Mode, which provides prioritised mitigation strategies aimed at helping an organisation mitigate cybersecurity incidents (ACSC, 2020: 2). The model highlights a number of key technical aspects that should be addressed for cybersecurity protection, including application whitelisting, patching of applications and operating systems, configuration of Microsoft Office macro settings, user application hardening, restriction of administrative privileges, multi-factor authentication and daily updates.

### **5.4 Develops a cybersecurity campaign**

Cybersecurity challenges will not be solved by technology alone; the human element is also one of the key elements in ensuring enhanced cyber resilience. Training and education could help staff identify and prevent cyberattacks. Additionally, (Microsoft, 2017: 13) proposes that cybersecurity awareness and training should be extended to citizens as well. A cybersecurity awareness and behavioural change campaign can encourage the public and small businesses to adopt simple behaviours to protect themselves against cyber threats. Developing cybersecurity campaigns for councillors and other local authorities would include seminars, conferences and exercises aimed at raising awareness on the security of their cyberspace.

### **5.5 Enforce cybersecurity culture**

A cybersecurity culture is necessary to inculcate acceptable user behaviour in cyberspace. Awareness and education are regarded as pillars in promoting a cybersecurity culture (Kortjan & Von Solms, 2014). Improving cybersecurity also requires managers to create and maintain cultures of cybersecurity within their local governments (Microsoft, 2017: 13). However, this should be conducted in cooperation with local elected officials, IT and cybersecurity staff, department managers, and citizens.

### **5.6 Prepare for cybersecurity incidents**

It is important to back up regularly, and test thoroughly for availability and integrity. However, the routine back up should also be extended to other devices used to render services and communication to citizens. Those devices include laptops, smartphones, tablets, etc. In the event of a cyber-attack, rapid response helps reduce the impact. Staff must be well versed in the incident response procedure. They need to know basics such as what procedures should be followed to contain the infection, the roles and responsibilities of all team members, and who should be notified.

## **6. Conclusion**

The growth of cyberspace will continue to advance thus security measures to enhance cyber reliance especially at local government level should be intensified. Similar to public safety, cybersecurity requires that the government implement good cybersecurity measures that will result in the protection and security of cybersecurity. Protecting communities is an important aspect of local government's role, alongside that of service delivery and creating more sustainable and cyber-resilient systems. To enhance economic growth, public safety, research and innovation around cybersecurity, public trust in the integrity of financial systems, information networks, and other critical information infrastructure systems managed by local government should be prioritised (Gagliardi et al., 2016). The paper highlights several challenges to achieving enhanced cybersecurity and those include that lack of funding, cybersecurity-related policy challenges, lack of cybercrime training by relevant law enforcement agencies, lower chance of being caught or prosecuted and poor public knowledge of cyber threats. Deloitte (2020: 3) states that emerging technologies such as automation and artificial intelligence can and should be used to augment an organisation's traditional cybersecurity efforts. However, such technologies do not eliminate the need for human experts. The said challenges, technologies and

the human factor in cybersecurity at the local government level all constitute fertile areas for further academic research.

## References

- Auditor General of South Africa. (2020). Consolidated general report on local government audit outcomes: 2010/2019. Available at: <https://www.agsa.co.za/Reporting/MFMAReports/MFMA2018-2019.aspx> [Accessed 20 January 2021].
- Australian Cyber Security Centre (ACSC). (2020). Essential Eight Maturity Model. Available at: <https://www.cyber.gov.au/sites/default/files/2020-06/PROTECT%20-%20Essential%20Eight%20Maturity%20Model%20%28June%202020%29.pdf> [Accessed 02 February 2021].
- BizNews. (2017). Biggest ever SA data breach: 60 million ID numbers leaked on real estate server. Available at: <https://www.biznews.com/global-citizen/2017/10/20/biggest-ever-sa-data-breach> [Accessed 19 July 2019].
- Calandro, E. (2018). Submission to the Inquiry into the role and responsibilities of the Independent Communications Authority of South Africa in Cybersecurity. Available at: <https://researchictafrica.net/wp/wp-content/uploads/2018/12/Submission-Cybersecurity-ICASA-vFINAL.pdf> [Accessed 22 January 2021].
- Coleman, Y. (2019). Cyber Accountability Model. Available at: <https://globalsmartcitiesalliance.org/?p=799> [Accessed 08 February 2021].
- Constitution of the Republic of South Africa (1996). Available at: <https://www.gov.za/documents/constitution-republic-south-africa-1996> [Accessed 10 February 2021].
- Deloitte. (2020). The changing faces of cybersecurity: closing the cyber risk gap. Available at: <https://www2.deloitte.com/content/dam/Deloitte/CA/Documents/risk/ca-cyber-talent-campaign-report-pov-aoda-en.PDF> [Accessed 10 September 2020].
- Deloitte. (2020b). The future of cyber. Available at: <https://www2.deloitte.com/global/en/pages/about-deloitte/articles/gx-future-of-cyber.html> [Accessed 10 February 2021].
- Duvenage, P.C., Jaquire, V. J. & von Solms, S.H. (2020). 'A Cyber Counterintelligence Matrix for Outsmarting Your Adversaries' in Journal of Information Warfare, 19(1): pp 1-11
- Eisenstein, L. (2020). Cybersecurity Standards for Local Government. Available at: <https://insights.diligent.com/cybersecurity-local-government/developing-cybersecurity-standards-local-government> [Accessed 16 October 2020].
- Gagliardi, F., Hankin, C., Gal-Ezer, J., McGettrick, A. and Meitern, M. (2016). Advancing Cybersecurity Research and Education in Europe. Available at: [https://www.acm.org/binaries/content/assets/publicpolicy/2016\\_euacm\\_cybersecurity\\_white\\_paper.pdf](https://www.acm.org/binaries/content/assets/publicpolicy/2016_euacm_cybersecurity_white_paper.pdf) [Accessed 01 May 2017].
- Gcaza, N. and Von Solms, R. (2017). A strategy for a cybersecurity culture: A South African perspective. The Electronic Journal of Information Systems in Developing Countries, 80(1), pp.1-17.
- Hadjizenonos, D. 2019. Municipal cybersecurity on a shoestring. Available at: <https://www.itweb.co.za/content/G98YdMLxNQRMX2PD> [Accessed 17 January 2021].
- Hubbard, J. (2019). SA business underplaying the danger of cybercrime? Available at: <https://www.fin24.com/Finweek/Business-and-economy/sa-business-underplaying-the-danger-of-cybercrime-20190313> [Accessed 17 April 2020].
- Institute of Directors of South Africa. (2018). Governing Body's Role in Cyber Resilience Corporate Governance Network. Available at: [https://cdn.ymaws.com/www.iodsa.co.za/resource/collection/05E93ACB-10BE-4507-9601-307A66F34BD8/CGN\\_Position\\_paper\\_13\\_Cyber\\_Resilience\\_April\\_2018.pdf](https://cdn.ymaws.com/www.iodsa.co.za/resource/collection/05E93ACB-10BE-4507-9601-307A66F34BD8/CGN_Position_paper_13_Cyber_Resilience_April_2018.pdf) [Accessed 18 April 2020].
- Masombuka, M. (2018). Towards an artificial intelligence framework to actively defend cyberspace in South Africa (Doctoral dissertation, Stellenbosch: Stellenbosch University). Available at: <http://scholar.sun.ac.za/handle/10019.1/105239>.
- Masombuka, M., Grobler, M. and Watson, B. (2018). Towards an Artificial Intelligence Framework to Actively Defend Cyberspace. In European Conference on Cyber Warfare and Security (pp. 589-XIII). Academic Conferences International Limited.
- Mabaso, N., J. (2018). Assessing the cyber-security status of the metropolitan municipalities in South Africa (Doctoral dissertation). Available at: <https://ukzn-dspace.ukzn.ac.za/handle/10413/18097> [Accessed 29 January 2021].
- Mcanyana, W., Brindley, C. and Seedat Y. (2020). Cyberthreat Landscape in South Africa. Available at: [https://www.accenture.com/\\_acnmedia/PDF-125/Accenture-Insight-Into-The-Threat-Landscape-Of-South-Africa-V5.pdf](https://www.accenture.com/_acnmedia/PDF-125/Accenture-Insight-Into-The-Threat-Landscape-Of-South-Africa-V5.pdf) [Accessed 09 January 2021].
- Merriam Webster (2021). local government. Available at: <https://www.merriam-webster.com/dictionary/local%20government> [Accessed 09 February 2021]
- Microsoft. (2017). Cybersecurity: Protecting Local Government Digital Resources. Available at: <https://icma.org/sites/default/files/18-038%20Cybersecurity-Report-hyperlinks-small-101617.pdf> [Accessed 016 June 2020]
- Nabe, C. (2020). Impact of COVID-19 on Cybersecurity. Available at: <https://www2.deloitte.com/ch/en/pages/risk/articles/impact-covid-cybersecurity.html> [Accessed 16 January 2021].
- Nicholas, P. J. and Pinter, J. (2017). Cyber resilience: digitally empowering cities. Available at: <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW6auc> [Accessed 02 December 2020].

***Mmalerato Masombuka, Marthie Grobler and Petrus Duvenage***

- Norris, D.F., Mateczun, L., Joshi, A. and Finin, T., (2020). Managing cybersecurity at the grassroots: Evidence from the first nationwide survey of local government cybersecurity. *Journal of Urban Affairs*, pp.1-23.
- Nyirenda-Jere, T. and Biru, T. (2015). Internet development and Internet governance in Africa. ISOC Report. Available at: <https://www.sbs.ox.ac.uk/cybersecuritycapacity/system/files/Internet%20development%20and%20Internet%20governance%20in%20Africa.pdf> [Accessed 12 March 2017].
- Osborn, E. and Simpson, A. (2015). November. Small-scale cybersecurity. In 2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing, pp. 247-252.
- Pandey, P., Golen, D., Peasley, S. and Kelkar, M. (2019). Making smart cities cyber secure: Ways to address distinct risks in an increasingly connected urban future. Available at: <https://www2.deloitte.com/us/en/insights/focus/smart-city/making-smart-cities-cyber-secure.html> [Accessed 20 January 2021].
- Smith, C. (2019). Major spike in SA cyberattacks, over 10 000 attempts a day - security company. Available at: <https://www.news24.com/fin24/Companies/ICT/major-spike-in-sa-cyber-attacks-over-10-000-attempts-a-day-security-company-20190429> [Accessed 03 December 2020].
- Sutherland, E. (2017). Governance of cybersecurity-The case of South Africa. *African Journal of Information and Communication*, 20, pp.83-112.
- Toulu, A. (2018). Cyber threats on African subjects. Available at: [https://www.ict.org.il/Article/2275/Cyber\\_threats\\_on\\_African\\_subjects#gsc.tab=0](https://www.ict.org.il/Article/2275/Cyber_threats_on_African_subjects#gsc.tab=0) [Accessed 20 August 2019].
- Thompson, L. N. (2019). Cybersecurity Best Practices for Municipalities. Available at: <https://www.nhmunicipal.org/town-city-article/cybersecurity-best-practices-municipalities>. [Accessed 02 December 2020].

# KSA for Digital Forensic First Responder: A job Analysis Approach

Ruhama Mohammed Zain, Zahri Yunos, Nur Farhana Hazwani, Lee Hwee Hsiung and Mustaffa Ahmad

CyberSecurity Malaysia, Cyberjaya, Malaysia

[ruhama@cybersecurity.my](mailto:ruhama@cybersecurity.my)

[zahri@cybersecurity.my](mailto:zahri@cybersecurity.my)

[farhana.hazwani@cybersecurity.my](mailto:farhana.hazwani@cybersecurity.my)

[hh.lee@cybersecurity.my](mailto:hh.lee@cybersecurity.my)

[mus@cybersecurity.my](mailto:mus@cybersecurity.my)

DOI: 10.34190/EWS.21.002

**Abstract:** The role of cybersecurity professionals is important in protecting the computer network in many organizations. Indeed, it is a big challenge to find highly skilled talents when it comes to specialized areas such as Digital Forensics First Responders (DFFR). We need more cybersecurity knowledgeable workers and experts, people with the right know-how to tackle rapidly evolving cyberattacks, risks and threats. The job analysis workshop in this study is an exploratory research method used to collect data through group interaction. This method provides an opportunity to observe the interaction among participants on the topic under study. This paper contributes to the findings of DFFR job analysis conducted by CyberSecurity Malaysia. Eight (8) participants took part in the job analysis workshop, comprising representatives from academia, government and industry. The findings summarize a set of Knowledge and Skills elements required to support a list of DFFR job tasks. The framework provides principal guidelines for training developers and end-users in developing training programs, with focus on competency-based assessment. This is useful for developing training courses and assessment questions towards DFFR certification. Some lessons learned from the job analysis process provide opportunities for improvement in future job analysis workshops.

**Keywords:** competency framework, cybersecurity professional, cybersecurity education, KSA

---

## 1. Introduction

There are numerous cybersecurity certifications in the world today. Some are geared towards management skill validation and some towards technical areas of cybersecurity. Professional certifications have become popular with employers as a way to distinguish capable applicants from the rest. Certifications are often used together with academic qualifications and work experience to judge whether a candidate has the necessary knowledge and skills for the job.

A professional certification must be valid in order to have value and very much signifies the capability and quality of the credential holder. "When we consider validity in the sense of assessment, we are considering whether or not the result of the assessment task actually achieves the purpose of setting the task. In discussing criterion referenced assessment, But et al (2007) note that "validity is at a premium, since assessment should be geared to showing whether a student can fulfill a criterion which the curriculum has been designed to enhance."

There are also detractors to the value of certifications in general. This is not helped by the fact that some certifications only measure routine memorization ability and not the actual job performance skill. To make matters worse, some so-called training and boot camps are designed to teach how to pass the certification exam. Such regimes produce what is lowly thought of as paper certification.

The Global Accredited Cybersecurity Education Certification (Global ACE Certification) introduced by CyberSecurity Malaysia, an agency under the Ministry of Communications and Multimedia Malaysia, is a holistic cybersecurity professional certification framework. The Global ACE Certification is developed in cooperation with government, industry and academia players. The goal of the Global ACE Certification is to create cybersecurity certifications that are world-class and truly reflect the ability of the certificate holder to perform the required tasks for a given job role. The Global ACE Certification is distinguished from the typical multiple-choice question certifications in that it requires examinations to combine knowledge testing and hands-on or skill-based testing. The objective is to evaluate whether candidates can actually perform the job in a practical setting.

## **2. Literature review**

It is noted that several studies have been conducted to assess cybersecurity certification programs. The studies are valuable to understand the perspectives of stakeholders regarding the perceived value and demand for cybersecurity certification programs and demand for certifications in general.

Caroll (2018) focused on developing an offensive/defensive cyberspace operations workforce. Twenty-three (23) participants took part in the study, including military, civilian and contractor cyberspace operations professionals in the US. The purpose of the study was to compare the KSA (Knowledge, Skills and Ability) as defined by the National Security Agency (NSA) for cyberspace operations fundamentals against the KSA perceived by the workshop participants. The respondents indicated that five core modules must be implemented: Networking (wired & wireless), Security Fundamental Principles, Telecommunications, Computer and Network Defense, and Vulnerability and Risk Management. The respondents also recommended several other modules to be considered in the program, but not mandatory.

Kornblum (2002) studied the training module on the preservation of digital evidence by First Responders. The paper describes the challenges First Responders face on the ground. The author proposed that First Responders must be careful not to destroy or taint digital evidence. The author also highlighted two major issues regarding digital evidence: transient data that is lost at shutdown and fragile data stored on hard disk that can be easily altered such as the last accessed time stamp. These recommendations can be used as a guideline in developing Digital Forensics First Responder training modules in a real-world environment.

A study to investigate the capabilities of cybersecurity and human workforce in response to the industrial control system (ICS) environment was conducted by Ani et al (2019). The study indicates that cybersecurity is an important element in achieving ceaseless industrial functions in a changing operational technology environment. The study also emphasizes that assessing the cybersecurity capability of the workforce is quite helpful in achieving an efficient workforce security consciousness.

But et al (2007) discussed real-world learning and real-world assessment tasks. A real-world learning experience is a practical task (real or simulated) that provides an opportunity for the practical application of theoretical knowledge. The assessment of real-world learning tasks needs to be process-based, with the task parameters having clearly defined learning outcomes and assessment criteria. In a real-world learning experience, it is crucial to define the purpose of the assessment and what aspects of the task are to be assessed. Real-world assessment tasks are valid in the sense that the application of a set of skills is required to solve a problem and this is what the curriculum is designed to enhance. The authors argued that the appropriateness of practical assessment versus theoretical knowledge assessment really depends on the tasks and discipline being studied.

The International Organization for Standardization (2012) developed ISO/IEC 27037 Information Technology-Security Techniques-Guidelines for Identification, Collection, Acquisition, and Preservation of Digital Evidence. The document provides guidelines for specific activities in the handling of digital evidence, which are identification, collection, acquisition and preservation of potential digital evidence that may be of evidential value. The ISO document can be used as reference in developing a KSA document.

## **3. Digital forensics first responder**

The cybersecurity field can be broken down into several major domains. This paper concerns one of the job roles in the digital forensic domain, namely the first responder. The certification domain must be clearly defined in order to “seek out, prioritize, select, and organize the most valuable knowledge, skill, ability, and dispositional components of real-world competence” (Zane, 2009). The Global ACE Certification uses a slightly modified nomenclature, which is the Digital Forensics First Responder (DFFR). In contrast, the ISO/IEC 27037 document refers to the first responder as a Digital Evidence First Responder (DEFRR).

According to ISO/IEC 27037 (2012), a Digital Evidence First Responder is an individual who is “trained and qualified to act first at an incident scene in performing digital evidence collection and acquisition with the responsibility for handling that evidence”. The idea is to collect data in a way that will not change the data from the original form. The reason is so the data can be trusted in the court of law when a case is prosecuted. It is also important for the DFFR to maintain a chain of custody record to demonstrate that the data was acquired, handled, transported and stored properly to maintain its integrity.

Digital evidence can be very easy to destroy due to its digital nature. Destruction can be caused by changes made to the evidence during acquisition, handling or examination. Care must be taken to minimize handling of the original devices and to maintain careful records of any changes or actions taken. One of the key principles of digital evidence is to “acquire the potential digital evidence in the least intrusive manner in order to avoid introducing changes where possible” (ISO/IEC, 2012). The most appropriate acquisition method must be selected, and the process documented in detail. The person acquiring the digital evidence must not overstep their level of competence.

It must be demonstrated that the digital evidence collected is relevant to the investigation due to the value of information it contains. The processes adopted when handling digital evidence must be auditable and repeatable. It is against this backdrop of digital evidence principles and requirements that the Knowledge, Skills and Attitudes of a Digital Forensics First Responder are developed based on the tasks typically performed by the person with this role.

#### **4. Importance of assessment**

The Global ACE Certification decided that assessment is based on real-world or on-the-job scenarios that emphasize practical skill testing. However, this type of assessment needs to consider the underlying “professional subject knowledge base and the extent to which this informs the task itself” (But et al, 2007).

To address this requirement, the Global ACE Certification decided to use the Knowledge, Skills and Attitudes (KSA) descriptor for each job role defined. The KSA elements map readily onto the Cognitive, Psychomotor and Affective domains of the Bloom taxonomy (Anderson and Krathwohl, 2001). Specifically, the Knowledge, Skills and Attitudes elements map onto the Cognitive, Psychomotor and Affective domains respectively. This makes the KSA descriptor suitable as a guide to develop both training content and assessment questions. The assessment is required in order to certify that the candidate can really do the job. Additionally, an assessment should be about verifying competence, which is defined as “the ability to perform the activities within an occupation” and “the ability to transfer skills and knowledge to new situations within the occupational area” (Kennedy et al, 2009).

In order for the assessment to be authentic, it must be ensured that it is possible to directly examine the candidate’s performance on tasks that require the application of knowledge and whether the candidate can craft justifiable answers or performance (Wiggins, 1990). This is addressed through a two-part assessment mechanism consisting of a straight-up multiple-choice examination and a hands-on practical assessment. This way, all three Bloom’s domains are assessed to determine how well the candidate has mastered the required competences in order to perform the job role tasks.

For a fair assessment, it is necessary to ensure that the assessment is carried out with multiple methods, such that valid inferences about the candidate’s ability can be made (McMillan, 2000). Again, this is addressed via the two-part assessment mechanism that exercises the whole spectrum of learning (and subsequently assessment) models (Anderson and Krathwohl, 2001). The Global ACE Certification is a professional certification that requires summative assessment. This is because the use of “summative assessment is generally to predict future performance, to license someone as competent, or as information for entrance to other academic institutions or for the selection boards of firms or professional bodies” (But et al, 2007).

This is in contrast to the formative assessment that may be more relevant in an academic setting. This type of assessment is more geared toward helping the candidate develop “by providing constructive feedback from which they can learn to identify the ways in which they need to improve” (But et al, 2007). It has been decided that formative assessment is not used to prevent training instructors from abusing the training events to become popular by virtue of having more trainees passing a certification exam.

### **5. Methodology**

#### **5.1 Participants**

The participants for the job analysis workshop were selected from the government, academia and industry sectors. They were identified as experts in the digital forensics field with years of experience. No replacement was allowed, so the right combination of knowledge and experience would be maintained for the workshop. A



total of eight participants comprising four external DFFR experts, three internal cybersecurity experts and one moderator attended the workshop to deliberate on the important knowledge and skills elements. The external experts comprised representatives from academia, government and industry.

## **5.2 Procedures**

The Global ACE Certification job analysis procedure was developed to produce the competencies required for a particular certification. The set of competencies comprises knowledge and skills that are required to perform a particular job. The job analysis procedure involves performing initial research, identifying the tasks for successful performance, identifying the required competence for each task, identifying the list of contents and weights, and lastly, identifying the assessment mechanisms.

A job analysis is done periodically or whenever there is any significant change in the profession. The information learned from a job analysis is used to determine whether to expand or reduce the scope of assessment for a particular certification. The contents or topics of a particular examination can then be outlined or updated. A panel that consists of a sufficient number of technical experts is appointed to serve in a job analysis study.

During the job analysis workshop, the participants were first asked to list the tasks required to successfully perform the job. The list was derived from existing documents, guidelines and standards currently in existence. The participants then provided additional input from other sources or based on their professional field experience. Each of the tasks was then given a ranking based on how important the task is to the job. Only tasks ranked "Important," "Very Important" or "Extremely Important" were kept for the next step.

The next step entailed listing the competencies required to perform each task from the previous step. This includes what one should know (knowledge) and what one should be able to do (skills) in order to execute each task. Once again, the participants were asked to draw upon their experience and existing body of knowledge to come up with the competencies. Similar competencies were consolidated to minimise the number of overlapping items. Then each competency was ranked according to importance for effective job performance.

## **5.3 Job analysis workshop**

The job analysis workshop was conducted to analyze and derive a set of tasks and the requisite knowledge and skills to do the tasks. A mix of experts from across the government, industry and academia areas were identified and invited to participate in the job analysis workshop. They were recognized digital forensics practitioners and experts from Malaysia. Prior to the workshop, each participant was introduced to the draft Knowledge, Skills and Attitudes KSA descriptor document for DFFR. They were also given an information packet containing the templates for listing and ranking the tasks required for a DFFR. The participants were asked to do some research on what the common tasks for a DFFR would be, then enter the tasks in the tasks template and bring it to the workshop.

On the day of the workshop, the facilitator first briefed the participants on the KSA descriptor structure and rationale, and asked for comments. Next, the participants were briefed on the job analysis process and how the workshop was going to be conducted. The participants were then presented with a proposed list of **tasks** that a person in the DFFR role would be expected to perform. This represents a raw list of inputs gathered by the workshop participants and from the National Initiative for Cybersecurity Education (NICE) framework (Newhouse et al, 2016). By the end of the workshop, the list will have been pruned and finalized based on the consolidated rankings of the task items given by the participants. The rankings served to reach consensus on what tasks are deemed important for digital forensics first responder. Appendix 1 shows the finalized list of tasks.

The participants were also given a proposed list of **knowledge** elements required by digital forensics first responder to fulfill a task. At the end of the workshop the list was finalized, such that only knowledge elements required to support the finalized tasks were maintained. This was done through a ranking process similar to that for the tasks, whereby consensus was reached by the experts. Appendix 2 shows the finalized list of knowledge elements.

Similarly, the participants were asked to go through a list of potential skills required for digital forensics first responder to fulfill a task. At the end of the workshop the list was finalized to maintain only the skills elements

required to support the finalized tasks. Again, this was done through a ranking process like that for the tasks, until the experts reached consensus. Appendix 3 shows the finalized list of skills.

It was decided that the Attitudes elements will not be discussed as a separate item because they were not going to be assessed specifically. Instead, the Attitudes elements will be blended into the training content by incorporating them into the case studies, group exercises and the overall training philosophy. This paper will not cover the Attitudes elements.

### 5.4 Ranking guide

In order to have the most important **tasks** elements appear in the finalized set of Digital Forensics First Responder **tasks**, only those that scored 3 or better (i.e. 3, 4 or 5) were maintained. Tasks that scored 2 or 1 were deemed not critical to the job role. The ranking scheme is shown in Table 1 together with the meaning behind each ranking score. The most important items scored 5 and the least important scored 1.

**Table 1:** Task ranking guide

Importance	Refer to either one of the frequency scales that is more relevant to the task	
	Frequency	Frequency
How important is this task to the job?	How often is the task performed?	Compared to all other tasks you perform, how much time do you spend performing this task?
0 = Not performed	0 = Not performed	0 = Not performed
1 = Not important	1 = Every few months to yearly	1 = Considerably less than most tasks
2 = Somewhat important	2 = Every few weeks to monthly	2 = Somewhat less than most tasks
3 = Important	3 = Every few days to weekly	3 = Same as most tasks
4 = Very important	4 = Every few hours to daily	4 = Somewhat more than most tasks
5 = Extremely important	5 = Hourly to many times each hour	5 = Considerably more than most tasks

Table 2 shows the **ranking scale for the Knowledge and Skills elements**. This will be the guide to determine whether a given **Knowledge or Skills element will appear in the finalized set**. Again, only the elements scored 3 or better (i.e. 3, 4 or 5) will be retained. Note that only the column labelled “Importance Scale” is considered. The other two columns (“Needed at Entry Scale” and “Distinguishing Value Scale”) are meant to guide hiring managers when considering candidates for employment and are not meant to decide whether to include or exclude the Knowledge and Skills elements from the final set.

**Table 2:** Ranking guide for knowledge and skills

Importance Scale	Need at Entry Scale	Distinguishing Value Scale
How important is this competency for effective job performance?	When is this competency needed for effective job performance?	How valuable is this competency for distinguishing superior from barely acceptable employees?
0 = Not Applicable	0 = Not Needed	0 = Not Applicable
1 = Not Important	1 = Needed the first day	1 = Not Valuable
2 = Somewhat Important	2 = Must be acquired within the first 3 months	2 = Somewhat Valuable
3 = Important	3 = Must be acquired with the first 4-6 months	3 = Valuable
4 = Very Important	4 = Must be acquired after the first 6 months	4 = Very Valuable
5 = Extremely Important	-----	5 = Extremely Valuable

## 6. Results

The results show that the workshop participants were in general agreement that a DFFR is only expected to perform data acquisition at a scene where an incident took place. The same DFFR is not expected to also do the data analysis. Although some participants did in fact include the analysis part into the tasks list, these were later

removed after clarifying that the analysis part is not included in the scope of the DFFR certification being developed.

The significance of making a forensically sound copy of the data is clearly seen from the output. This is important especially when the investigation result depends on accurate and undisputed evidence. The final set of KSA descriptors for DFFR contains 20 Tasks elements, 28 Knowledge elements and 12 Skills elements. A matrix of the Tasks, Knowledge and Skills together with their importance rankings is shown in Table 3.

The following table shows the finalized set of competencies proposed by the workshop attendees. The required competencies are mapped to each task determined to be important for digital forensics first responder to deliver the job. For a full description of each task (e.g. T3), knowledge (e.g. K5) and skill (e.g. S2), please consult Appendices 1, 2, and 3 respectively.

**Table 3:** Critical tasks and competencies linkage

Task	Knowledge	Skills
T1	K5, K7, K8, K10, K11, K15, K16, K20, K21, K22, K25, K27, K38	S1, S2, S3, S5, S8, S11, S14
T2	K5, K7, K8, K10, K11, K15, K16, K27, K30, K38, K39	S18, S21
T3	K5, K7, K8, K10, K11, K15, K16, K20	S3, S8, S14
T4	K5, K8, K10, K11, K15, K16, K25, K29, K30, K35, K37, K38, K40, K41, K42, K43, K46, K47, K50, K51	S7, S12, S14, S21, S24
T5	K25	S3
T6	K25, K27, K38	S2, S3, S7, S8
T7	K15, K16, K37, K38	S11, S21, S24
T8	K20, K22, K37, K38, K46	S2, S7, S11, S21
T9	K16, K19, K20, K22, K27, K29, K35, K37, K38, K39, K41, K42	-
T10	K5, K7, K10, K11, K15, K30, K37, K38, K45, K46	-
T11	K15, K16, K19, K20, K22, K25, K30, K37, K38, K41, K42, K43	S2, S7
T12	K16, K22, K25, K27, K29, K37, K38	S3
T13	K16, K22, K27	S3, S8
T14	K16, K22, K27	S1, S2, S8
T15	K16, K19, K22, K30, K37, K38, K41, K42, K43	-
T16	K7, K10, K11, K16, K19, K22, K25, K27, K37, K38	-
T17	K5, K7, K8, K10, K11, K16, K25, K39	-
T18	K5, K7, K10, K11	-
T19	K5, K7, K8, K10, K11, K38, K39	-
T20	K5, K7, K8, K10, K11, K16, K27, K38, K39	S8

## 7. Discussion

From the initial list of 52 tasks given to the experts, a total of 32 tasks were dropped after the experts reached consensus. A closer examination of the dropped tasks indicates that they are either not directly relevant to digital forensics first responders, or they concern the analysis part of digital forensic work or the pre-analysis tasks required before analysis is possible. This is in line with the design decision to exclude analysis tasks from the digital forensics first responder job requirement.

From the initial list of 51 knowledge elements given to the experts, a total of 23 elements were dropped from the final list. A closer inspection of the dropped knowledge elements reveals that some entail rather basic knowledge deemed to have been mastered by the digital forensics first responder and therefore understood to be already covered in other curricula. This is to avoid making the digital forensics first responder training, which is an intermediate level training, too lengthy as it has to cover basic material.

From the initial list of 28 skills elements given to the experts, a total of 16 elements were dropped from the final list. A closer inspection of the dropped skills elements reveals that some are foundational skills supposed to have been acquired by the digital forensics first responder and therefore understood to have already been covered in other curricula. Again, this is to avoid making the intermediate-level digital forensics first responder

training too lengthy, because it has to cover the basic skills elements. Other elements are basic computer and networking skills that are specified as training prerequisites.

It is acknowledged that no single assessment can ever hope to measure a candidate’s competence with one hundred percent accuracy. The assessment is also not a guarantee of future on the job performance. However, it is argued that the assessment is still useful if it covers the knowledge and skills required to perform the most important tasks. In order to achieve this, the finalized set of tasks are ranked according to their importance. The practicum part of the assessment is then crafted to cover at least the top ten most important tasks.

It is also important to keep track of the job performance of a certified person. In terms of validation, the Global ACE Certification aims to engage with and get support from employers to respond to future surveys on the effectiveness of the certification. This is a way to validate the tasks, the mapping of the tasks to knowledge and skills, and the assessment methods.

## 8. Conclusion

This study contributes findings from the job analysis workshop conducted for the DFFR role. The proposed competencies can be used by training developers to develop training programs that focus on competency-based assessment. For future work, a survey of how effective the assessment is in predicting actual job performance will be reported. The same job analysis methodology will be used for other cybersecurity job roles by applying lessons learned from this workshop and involving a much larger expert panel to get a better representative consensus.

### Appendix 1: Digital forensics first responder tasks

JAT#	Task
T1	Create a forensically sound duplicate of the evidence (e.g., forensic image) that ensures the original evidence is not unintentionally modified, to use in the data recovery and analysis processes. This includes but is not limited to hard drives, floppy diskettes, CDs, PDAs, mobile phones, GPS, and all tape formats.
T2	Provide a technical summary of findings in accordance with established reporting procedures.
T3	Ensure the chain of custody is followed for all digital media acquired in accordance with the Federal Rules of Evidence.
T4	Identify digital evidence for examination and analysis in such a way as to avoid unintentional alteration.
T5	Perform hash comparison against an established database.
T6	Prepare digital media for imaging by ensuring data integrity (e.g., write blockers in accordance with standard operating procedures).
T7	Recognize and accurately report forensic artifacts indicative of a particular operating system.
T8	Extract data using data carving techniques (e.g., Forensic Tool Kit [FTK], Foremost).
T9	Utilize a deployable forensics toolkit to support operations as necessary.
T10	Understand the scenario and derive a scope of work.
T11	Carry out a thorough search for artifacts that may contain digital evidence.
T12	Identify and secure digital devices for evidence acquisition.
T13	Identify, collect, and seize documents or physical evidence to include digital media and logs to avoid unintentional alteration.
T14	Collect devices that potentially contain digital evidence.
T15	Use specialized equipment and techniques to catalog, document, extract, collect, package, and preserve digital evidence.
T16	Initiate and maintain the preservation of evidence throughout digital evidence handling.
T17	Gather and preserve evidence used in the prosecution of computer crimes.
T18	Label and tag evidence for transport.
T19	Document the original condition of digital and/or associated evidence (e.g., via digital photographs, written reports, hash function checking).
T20	Follow the SOP from the crime scene to the lab.

### Appendix 2: Digital forensics first responder knowledge elements

JAC#	Competency
K5	Understanding of legal requirements for digital evidence.

JAC#	Competency
K7	Knowledge of laws, regulations, policies, and ethics as they relate to cybersecurity and privacy.
K8	Knowledge of legal governance related to admissibility (e.g. Rules of Evidence).
K10	Knowledge of electronic evidence law.
K11	Knowledge of legal rules of evidence and court procedure.
K15	Knowledge of investigative implications of hardware, operating systems, and network technologies.
K16	Knowledge of types and collection of persistent data.
K19	Knowledge of deployable forensics.
K20	Knowledge of data carving tools and techniques (e.g., Foremost).
K21	Knowledge of anti-forensics tactics, techniques, and procedures.
K22	Knowledge of forensics lab design configuration and support applications (e.g., VMWare, Wireshark).
K25	Knowledge of encryption algorithms.
K27	Knowledge of incident response and handling methodologies.
K29	Knowledge of security event correlation tools.
K30	Knowledge of network security architecture concepts including topology, protocols, components, and principles (e.g., application of defense-in-depth).
K35	Knowledge of debugging procedures and tools.
K37	Knowledge of system administration, network, and operating system hardening techniques.
K38	Knowledge of system administration concepts for operating systems, such as but not limited to, Unix/Linux, IOS, Android, and Windows.
K39	Knowledge and understanding of operational design.
K40	Knowledge of reverse engineering concepts.
K41	Knowledge of malware analysis tools (e.g., Oily Debug, Ida Pro).
K42	Knowledge of malware with virtual machine detection (e.g. virtual aware malware, debugger aware malware, and unpacked malware that looks for VM-related strings in the computer's display device).
K43	Knowledge of binary analysis.
K45	Knowledge of specific operational impacts of cybersecurity lapses.
K46	Knowledge of system and application security threats and vulnerabilities (e.g., buffer overflow, mobile code, cross-site scripting, Procedural Language/Structured Query Language [PL/SQL] and injections, race conditions, covert channel, replay, return-oriented attacks, malicious code).
K47	Knowledge of hacking methodologies.
K50	Knowledge of application security risks (e.g. Open Web Application Security Project Top 10 list).
K51	Knowledge of web mail collection, searching/analyzing techniques, tools, and cookies.

### Appendix 3: Digital forensics first responder skills elements

JAC#	Competency
S1	Is able to physically disassemble digital devices.
S2	Is able to conduct acquisition from digital devices.
S3	Is able to preserve digital evidence.
S5	Skill in developing, testing, and implementing network infrastructure contingency and recovery plans.
S7	Skill in identifying and extracting data of forensic interest in diverse media (e.g., media forensics).
S8	Skill in collecting, processing, packaging, transporting, and storing electronic evidence to avoid alteration, loss, physical damage, or destruction of data.
S11	Skill in conducting forensic analyses in multiple operating system environments (e.g., mobile device systems).
S12	Skill in analyzing anomalous code as malicious or benign.
S14	Skill in processing digital evidence, to include protecting and making legally sound copies of evidence.
S18	Skill in interpreting debugger results to ascertain tactics, techniques, and procedures.
S21	Skill in identifying, modifying, and manipulating applicable system components within Windows, Unix, or Linux (e.g., passwords, user accounts, files).
S24	Skill in using virtual machines. (e.g., Microsoft Hyper-V, VMWare vSphere, Citrix XenDesktop/Server, Amazon Elastic Compute Cloud, etc.).

## References

- Anderson, L. and Krathwohl (2001) *A Taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*, Longman.
- Ani, U.D. et al (2019) "Human factor security: evaluating the cybersecurity capacity of the industrial workforce," *J. Syst. Inf. Technol.*, vol. 21, no. 1, pp. 2–35.
- But, J., Fleckhammer, L., Oates, G., and Rickards, H. (2007) "Assessing real-world learning experiences validly and reliably."
- Carroll, J. (2018) "Offensive and Defensive Cyberspace Operations Training: Are we There yet?," *Eur. Conf. Cyber Warf. Secur.*, pp. 77–86.
- International Organization for Standardization (2012) "ISO/IEC 27037:2012 Information technology —Security techniques —Guidelines for identification, collection, acquisition, and preservation of digital evidence (ISO No. 27037:2012)." International Organization for Standardization.
- Kennedy, D., Hyland, A. and Ryan, N. (2009) "Learning Outcomes and Competences, Bologna Handbook," *Introd. Bol. Object. Tools*, no. B 2.3-3, pp. 1–18.
- Kornblum, J. (2002) "Preservation of Fragile Digital Evidence by First Responders," *Digit. Forensics Res. Work. (Vol. 8)*, pp. 1–11
- McMillan, J.H. (2000) "Fundamental assessment principles for teachers and school administrators.," *Pract. Assessment, Res. Eval.*, vol. 7, no. 8, pp. 89–103.
- Newhouse, W., Keith, S., Scribner, B. and Witte, G. (2016) "National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework," *NIST Spec. Publ. 800-181*.
- Wiggins, G. (1990) "The case for authentic assessment. - practical assessment, research & evaluation," *Pract. Assessment, Res. Eval.*, vol. 2, no. 2, pp. 1–3.
- Zane, T.W. (2009) "Performance assessment design principles gleaned from constructivist learning theory (Part 1)," *TechTrends*, vol. 53, no. 1, pp. 81–90.

# The Unrehearsed Boom in Education Automation, Amid COVID-19 Flouts, a Potential Academic Integrity Cyber Risks (AICR)!

Fredrick Ochieng' Omogah

Medical Informatics, I.T & Computer Sciences at the Uzima University, Kisumu, Kenya

[fo2001ke@yahoo.com](mailto:fo2001ke@yahoo.com)

[fomogah@gmail.com](mailto:fomogah@gmail.com)

DOI: 10.34190/EWS.21.090

**Abstract:** Covid-19, a Severe Acute Respiratory Syndrome SARS-CoV-2, is an aggressive and infectious disease responsible for massive health havoc and resulting in high mortality rates globally. As a result, on 11th March 2020, the WHO declared Covid-19 a world pandemic due to its grievous impact on human health and livelihood. Learning and teaching in institutions have been disrupted following lockdowns and subsequent closures of all learning institutions across the globe. This pandemic has been quite surging. Many renowned professors and doctors in Kenyan and African academia have perished. Kenya and Africa are left with no choice in education but to use online platforms. The unrehearsed boom in education automation by universities may be a potential academic integrity cyber risk because this rush is more than anticipated. Even though the pandemic could be a wake-up call, industry players and stakeholders should re-design the education sector to be compatible with the emerging digital economies and globalized villages we currently live in. As a challenge during the 21st Century, Covid-19 has forced many organizations to shift their day-to-day activities to rely more on technology, and our universities have not been left behind. One may ask, "Under what circumstances is the shift to and reliance on technology taking place?" Covid-19 could be a silver lining for education which is a great idea; however, education automation should NOT only be focused on the pandemic and how well technology can be used as a new "normal" but also how bad things can get in the event of technology failures and potential criminal conducts. Technology alone can never be a solution to automation. Better approaches MUST include People, Process then technology (PPT) so that a formal way for aligning technology with education strategies can be achieved to nurture best practices and controls for successful education automation implementation.

**Keywords:** impact on human health, online learning platforms, education automation, academic integrity cyber risks, learning management system, unrehearsed education

---

## 1. Introduction

The distressing reports on Covid-19's impulsive outbreak in Wuhan (China) and its surging global spreads have brought a paradigm shift in businesses following massive disruptions worldwide. Since January 2020, many economies have been sinking with high unemployment rates, brewing far-reaching emotional complications and loss of lives due to the Covid-19 pandemic.

Achieving competitive advantage in today's businesses is determined by the effectiveness with which an organization procures, governs, and uses its I.T. infrastructure. Information technology infrastructure is a backbone and an enabler in the new industrial changes. Before Covid-19, business interactions had been shifting gradually to online platforms. The pace at which I.T is supporting this venture is also on the increase. Universities and middle-level colleges, where I.T. knowledge, expertise, and research capacities are hosted have been striving to convert I.T investments into tangible business outcomes for other firms.

The Internet, the World Wide Web (www), and Internet of Things (IoT) devices are now connecting students and lecturers in ways that challenge the very conventional education dispensation concept. As a result, the higher education ecosystem's physical place is diminishing due to more online interactions.

Education is a critical sector for a future stable economy. Failure to strike a balance in education automation to help coordinate various information systems and technology needs of each level in academic processes will open flood gates to online enemies. The purpose of the study was, therefore, to help in the successful implementation of online teaching and learning to deliver value education outcomes during the Covid-19 pandemic using an online platform.

## 2. State of technology in the Kenyan higher education sector

The advent of the Covid-19 pandemic has been an all-inclusive change strategy and a paradigm shift. As an attempt to mitigate disrupted face-to-face education, it has brought some returns on investment for the education sector in Kenya. However, it's worth noting that such innovation arrangements also come with high

risks and chances of failure. In the wake of the Covid-19 pandemic, the Presidential executive directive led to the closure of all learning institutions in Kenya. Learning and teaching modalities changed to online as was advised by the Ministry of Education (M.o.E), that online learning be implemented to mitigate Covid-19 disrupted education activities. Before the Covid-19 pandemic, quality teaching and learning deliverables using online platform were perused mainly to leverage technology to address deficiencies in universities.

Enterprise Resource Planning (ERP) systems that enhance communications and consolidation of data for universities is now a focus by the Commission for Higher Education (CUE) in Kenya. This has been made mandatory, and has put many institutions on their toes towards automation. Additionally, the resources of production within universities will always remain resources and never become production if there are no relevant personnel, processes, and technology to change them to a product.

Departmental/faculties' or schools' activities within a given university can only achieve measurable objectives and deliverables through the process of management by skilled university personnel. Currently, we are living during the information age, where personnel are highly involved in knowledge creation. Information is so vital and must be communicated to all departments. Technology is used to enable the connection of various departments and also to pass needed information instantaneously within an organization. Just as the human heart pumps oxygenated blood all over the human body; the Enterprise Resource Planning (ERP) Management Information System (M.I.S.) does the same for information in organizations. Shared information across (ERP) and (L.M.S.) may be exposed to threats and numerous dangers because it can be misused or abused.

## **2.1 The challenges to successful implementation of online education platforms**

Technology use embodies both opportunities and also sources of defenselessness in businesses. Today's stakeholders are reluctant in carrying out performance measurement on I.T. functions for governance strategy to ensure optimization and alignments to core business objectives. There is a big deficiency when focusing on key I.T. investments by stakeholders coupled with underdeveloped I.C.T. Infrastructures and challenges in personnel capacities to build and handle online platforms.

### *2.1.1 Power interruption*

Power outages have been affecting the smooth running of online learning activities in our local universities. The Learning Management System (L.M.S.) always has to be restarted all the time whenever the power goes off. This happens due to the ineffective and absence of standby generators in many institutions. Other energy sources like solar and windmills are not widely used and are insufficient. The unavailability of Uninterruptible Power Supply (U.P.S.) equipment on campuses has been a disappointment and resulted in undesirable situations during online sessions. Whenever an online examination is in session, the process can be interrupted, making both faculty and students panic more during online assessments.

### *2.1.2 Inadequate internet connectivity*

Difficulties have been experienced because of challenges in the internet connectivity resources. Currently, taking keen interest one will wonder why there are more fluctuations in internet connectivity than before, majorly being experienced in Kenya and other parts of Sub-Saharan Africa. Because of Covid-19, more face to-face activities have been shifted to rely more on the use of the Internet to work at home and so more users than before are on underdeveloped infrastructures for connectivity supply. This has put most of the learning institutions on shared service internet connectivity resources with dynamic I.P. addresses as opposed to a dedicated resource. The dedicated connectivity resource has a static Internet Protocol (I.P), which provides stable internet connectivity; a service which is still out of reach for many struggling institutions. The reason for connectivity problems can be pegged on the heavy taxation imposed on providers by the taxman. Online education is very susceptible to Internet connectivity fluctuations. I.T personnel always have to juggle around during internet downtime, resetting, and even reconfiguring available back-up routers during the unfortunate event as a business continuity plan. Due to over reliant and high uptake of technology, online presence is more than before, since the demand for Internet connectivity has increased. This phenomenon accelerated by the Covid-19 pandemic has created very unusual connectivity traffic because Internet users have increased in numbers more than before. It is also important to note that the increase in Internet based business applications has also been a major cause of cybercrime during the Covid-19 pandemic. With this, attackers take advantage to flood organizations' network infrastructures with spoofed packets from unknown sources. This led to the



organizations' bandwidth exhaustion as both users and criminals struggle over available bandwidth. Paralyzing online activities especially education services online. The fight for bandwidth can slow down the traffic for long hours during a day's work.

### *2.1.3 High tax rates*

Taxes imposed on I.T. equipment and connectivity bandwidth is a hindrance to online education. With this, faculty and students are unable to embrace online interactive activities. Financial matters have worsened, especially in private and upcoming institutions of higher learning. The cost of connectivity and associated I.T. equipment are still out of reach for common consumers. More ever, our universities have been languishing in the eras of declining capitations and reduced funding. Covid-19 impacts negatively on universities' finances because they get sustained from fees paid by students who are also worried about another round of tuition increases. Lockdowns and closures hindered fee payment because students have been at home and, therefore, are not paying fees as usual. Many universities became bankrupt and have not been able to procure relevant I.C.T. equipment that meets the threshold for setting up online platforms to mitigate disrupted learning and teaching.

### *2.1.4 Assumption that technology will happen by itself*

Faculty and administration personnel in our universities by default are hardwired. It has since brought many organizations to their knee at the mid-flight during automation attempts; the fundamental knowledge that technology is always a process and not just a product; and that it does not happen by itself is a safer playing ground for focused organizations. There should be no fence-sitters in attempts to implement and execute I.T. projects, which aim to bring change in organizations. Many attempts to launch online education platforms in Sub-Saharan Africa suffered this assumption. Planners and Policymakers, many a time, go overboard with squid priorities, finding it difficult to draw a comprehensive budget that would fully support the growth of I.T. infrastructures; in the long run, the I.T. sector is ignored and comes last in their wish list. Covid-19 is still here with all learning institutions, and we are not sure when it will go away! What are they doing about it so that I.T. adoption during this pandemic can remain silver lining and "new normal as they say it?"

### *2.1.5 Overwhelmed faculty members*

The academic staff has arrays of proficient obligations on their shoulders, ranging from researches, teaching, assessing, among many other duties as may be assigned in addition to their faculty demands. All these cannibalize online teaching and learning activities. As a result, incidences similar to academic misconducts may go unnoticed because of workload, as explained by Morris and Carroll (2016) and that cases of academic misconduct have been overlooked in the U.S. because they are usually viewed as "minor" offenses that would eventually be penalized by someone else. There is a likelihood of academic misconducts being overlooked in online remote delivery and assessment during the Covid-19 pandemic when faculty members are over-occupied.

## **3. Online quality education loop-holes**

Information resources in line with the online academic platform are very sensitive and subject to abuse by students, community, faculty members, and outsiders. When this system is porous, finances and academic grades may not maintain the integrity they deserve. A lot of caution should be taken to restrain ill motives.

### *3.1.1 Lack of training in the online platform*

Covid-19 pandemic caught everyone off-guard. Online education was mainly conducted to address learning deficiencies in our universities. Automation has been at a very slow pace and not keenly perused by most of our local universities. The mounting pressures to mitigate disrupted face-to-face education did not give faculty members and students the space or time to prepare. Handling online platforms in teaching, examination setting, and assessments during the Covid-19 pandemic has been a nightmare because of many assumptions that faculty members and their students are computer literate. There is a serious skill gap in online strategies for handling university education by faculty members to enhance quality, relevance, and students' experience.

### *3.1.2 Lack of technology alignment to education strategy*

It is believed that the use of technology should bring quality education delivery and nothing more or less. In order to achieve this strategy, technology MUST be managed and governed to provide a framework of Best

Practices and Controls for successful implementations. Alignment of technology to the core business in education strategy should deliver measurable goals if embraced. It is the order of the day that policymakers never prioritize this alignment in order to draw a comprehensive budget that would fully support the growth of I.T. infrastructures to support education online.

### *3.1.3 Missing online academic integrity policy*

Many African higher learning institutions are still in limbo when maintaining and addressing gaps in academic integrity for online students' assessment. It is still a journey for many developing counties to familiarize themselves with what it takes, not only to be able to carry out invigilation and assessment electronically, which needs careful scrutiny of online algorithms, but first of all, just to launch content modularization and e-platform mounting and to keep abreast with the development of academic integrity policy of the same.

### *3.1.4 Incompatible I.C.T. equipment*

Since the Covid-19 pandemic, it has been all systems go to struggle and pursue online learning and teaching to mitigate disrupted face-to-face conventional education processes. Lack of knowledge about relevant I.C.T. equipment to handle online education and unpreparedness has led to even more challenges than expected during this pandemic. The majority of faculty members and students can only afford very old and low-capacity computers, while in some instances; the use of smart phones poses many challenges, especially during online examinations. For instance, during assessment sessions, a candidate using a mobile phone may receive a phone call; this would disrupt his/her session leading to a restart of examination all over again. Since the online examination is timed, this incident will be inactivity which can be translated into an irregularity for some L.M.S. online examination rules.

### *3.1.5 Change resistance*

Negative attitudes and behavior by some faculty members can make students resist online sessions. This challenges both faculty and students. The idea of online education seems to have been left for I.T. departments alone because both the faculty and student community ignore user training or never pay attention to opportune user training on handling online learning and teaching activities. Because of this, the whole process can be abandoned if the I.T. team is not resilient enough. For good progression and success, user training will have to be repeated all over again. Many students never follow instructions, especially during online assessments, because they want shortcuts; that leads them to examinations' irregularity and subsequently getting logged out of the exam session without a successful submission.

### *3.1.6 Practical sessions and assessments challenge*

Laboratory activities for Computer Sciences/Information Technology, Engineering, and Health Science programs that require face-to-face interactions have become a challenge. Practical sessions had to be suspended because Covid-19 spreads through direct human contacts, environments' surfaces, and objects handled by infected persons (WHO, 2020). This to date has delayed assessments, and some curriculums have not been covered within the stipulated time. Additionally, conducting practical such as ward rounds for medical schools requires very expensive systems such as telemedicine, which many African universities are yet to acquire. Arguably, when practical laboratory procedures are conducted online, it's perceived as not meeting the required threshold by some quarters locally.

## **4. Materials and methods**

A cross-sectional research design was conducted using quantitative and qualitative research in which investigators administer a survey to a sample on a score scale between 0 and 10 (quantitative data) from a cross-section of two local universities' faculty members and students whose universities attempted online learning and teaching in the time of Covid-19 pandemic between March and December 2020. The survey described their attitude, and opinions, beliefs and characteristics (qualitative data) using e-mailed questionnaires which statistically analyzed the data to describe trends about responses on questions and tested research questions

### *4.1.1 Data variable*

The study sought to identify the relationship between performance measurement, academic integrity, faculty and students' capacities versus education automation.

4.1.2 Results

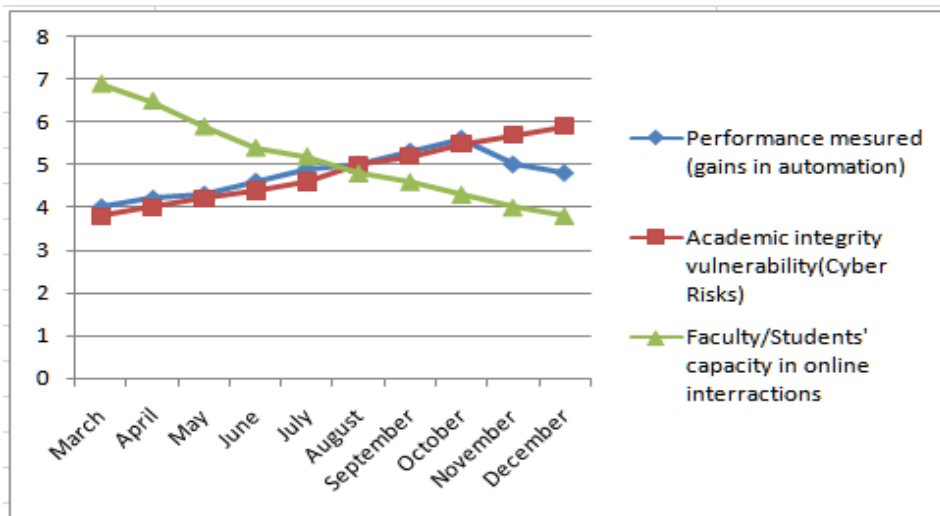
The figure below presents a paradigm shift with gains and risks scored on a scale between 0 and 10 from a cross-sectional survey study in two Universities reflecting experiences and expressions of attitudes, opinions, beliefs, and characteristics with the saddened change into online education amid the Covid-19 pandemic between March and December 2020. Performance of the paradigm shift has been felt with some gains from online education dispensation, salvaging disrupted face-to-face learning, and teaching activities by few universities.

**Table1:** Discoverable data on a score scale between 0 - 10 on the unrehearsed education automation by some Kenya universities during covid-19 pandemic between March and December 2020

Online attempts in 2020	Performance mesured (gains in automation)	Academic integrity vulnerability(Cyber Risks)	Faculty/Students' capacity in online interractions
March	4	3.8	6.9
April	4.2	4	6.5
May	4.3	4.2	5.9
June	4.6	4.4	5.4
July	4.9	4.6	5.2
August	5	5	4.8
September	5.3	5.2	4.6
October	5.6	5.5	4.3
November	5	5.7	4
December	4.8	5.9	3.8

4.1.3 Data analysis

According to analyzed data, online academic integrity remains a challenge because many faculties are still handicapped in handling security challenges in online platforms. Performance measurement progressively showed some gains during online learning; however, when proceeding to online assessments, a decline is evident at this point; most students would want to take advantage of shortcuts when being assessed online and faculty capacities in handling online assessments started to diminish. Because all these activities MUST take place online, it is likely to get messy with the current trends in technology where personnel, faculty members, and students are still incapacitated due to shifting work modalities for continuity in education and the subsequent emergence of recent cyber threats incidences amid covid-19 pandemic.



**Figure 1:** Unrehearsed education automation by Kenyan universities during COVID-19 pandemic between March and December 2020 with gains and potential online academic integrity and cyber risks

## **5. Conclusion**

Presently, automation and rationalization are at the pick of the desired change in our universities. However, they are slowly changing and moving strategies; they are expected to bring about modest returns on investments for the education sector and little academic risk. The Covid-19 pandemic has fueled this change strategy into an unexpected paradigm shift that has brought some greater returns from automation. It happened so fast, with many stakeholders in education taking risks to automation education without enough preparation. Stringent security and privacy consideration should be in place when designing and building systems for our society if we are sure we do not want future disappointments. Remember, just as the devil and sin are always ahead of us, so do cybercriminals to education innovations.

## **6. Recommendations**

Currently, most wide spread organized online criminal conducts are hacking, identity theft, Denial of Service (DOS) attacks, ransom ware attacks, phishing, computer fraud, copyright violations, child pornography and social engineering among others. Many of these crimes initiated through insider threats and can be directed to online education during Covid-19 pandemic.

The absence of security to online education platforms' resources breeds security challenges. The education sector is very critical in any economy globally. Stakeholders and regulators MUST be part of the team players in this online strategy to look into specific measures for mitigating the impact of the global lockdowns in education institutions by using online platforms. This will enhance activities and provide security measures to ensure a good level of integrity in learning and assessments and ensure preparedness in coping with future disasters and pandemics.

### *6.1.1 Calls to manage risks*

Being proactive through risk management enables the analysis of the likelihood of potential cyber-attacks to I.T. infrastructure, where a Learning Management System (L.M.S.) is hosted. Risk management is an assurance to institutions that security measures would be in place to guard against incidences and threat attempts. Activities like threat detections and reactions to illegal or unusual entries being made into the system are better ways to confirm intrusions to give real-time responses to incidences in a controlled manner. Measures like Intruder Detection Systems (I.D.S.), callback modems, or a controlled system to trigger the whole system's temporal paralysis as security mechanisms are a good direction for managing risks. Well-managed risks are a fallback for disaster recovery and business continuity. Timely reporting of threats or attack incidences and awareness creations are among the key undertakings to pursue.

### *6.1.2 Creation of I.T. security design team*

The design team at the infrastructural level in any university would form a representation from a cross-section of employees. This will affirm conformity with the policies and procedures that the team comes up with as corporate I.T. governance. This team would compose security protocols to assess and analyze security design features within I.T. infrastructure and Learning Management Systems (L.M.S.)'s modules to ascertain its threshold and filter out all potential criminal conduct.

### *6.1.3 "Zero rate" or "exempt" taxes on I.T. equipment*

A major hindrance to online education has been high taxes imposed on I.T. equipment and services by governments in Sub-Saharan Africa. This is also why many institutions are unable to launch online platforms to date to mitigate Covid-19 disrupted face-to-face learning. Expensive I.T. equipment, Internet connectivity bandwidth, and exploitation by I.T vendors is very discouraging. If government policymakers really want to support education during Covid-19, they should apply preferential rates to goods and services that support online education. Better options are making I.T equipment either "zero rated" or "exempt" to support education during this pandemic.

### *6.1.4 University Funds Board intervention*

University funds board (U.F.B.) should help universities mitigate the effect of the post-Covid-19 pandemic. This would ensure sustainable financial help to support the required online education resources for students and

faculty members. This would be a very strategic attempt during the Covid-19 pandemic and other unforeseen future pandemics to circumvent unnecessary breaks in academic flows.

#### ***6.1.5 I.T. governance strategy***

It is believed that technology should bring quality education delivery and nothing more or less. To achieve this strategy, technology MUST be governed to provide a framework for Best Practices and Controls resulting in successful implementations. With this, we are sure of the alignment of technology to core education as a business strategy, after which technology is expected to deliver.

As a major step, hands-on, through I.T. governance, services to faculty and students should be secured. Through the Commission for University Education (CUE) regulations, higher learning institutions should create guidelines and standards on how I.T. infrastructures are developed and guarded to deliver education with reduced potential incidental pitfalls that may be sneaked into online education activities. I.T. governance puts risk management as a tool in place to have trusted systems for education service delivery. I.T. governance brings assurance for secure communications across the organization's I.T. infrastructure.

Security measures are tailored to guard against cyber threats because they identify critical and vulnerable assets in online education platforms. Online academic integrity can only be maintained if devices in the infrastructure are secure and if faculty members' and students' conduct would be aligned to confidentiality, integrity, and availability in the entire arrangement. Also, education delivery services can only depend on this infrastructure if all technologies involved would behave as expected; otherwise, the entire business can just be compared to a ticking time bomb.

#### ***6.1.6 Timely reporting of online threats incidences***

Timely reporting of attack incidences and raising awareness/training are vital undertakings. For quality and integrity, accurate time tracking of online activities on the Learning Management System (L.M.S.) and entire network infrastructure would be key security measures to adopt to ascertain if the learner is the actual student. It is strategic to know and understand the enemies of online education platforms: Who are they? What are their motives? How do they get in? What do they do after they are in? "Once we know our weaknesses, the attackers will cease to do us any harm", says Georg Christoph Lichtenberg, German Physicist, and Philosopher (1742–1799)

#### ***6.1.7 Develop strategic partnerships and networks among stakeholders***

Stake holder's engagement should be constituted as policymakers, think tanks, scholars, I.T. vendors, and regulators for university education. This will maintain continuity in I.T. security and management strategy, which will also trickle down to the entire I.T. infrastructure management to guard the security mechanism around the Learning Management System (L.M.S.) so that it is not compromised. Additionally, regulating online education by the Commission for University Education (CUE) will help administer what is being offered online and evaluate if the threshold is met.

#### ***6.1.8 Strategic online assessment for academic integrity***

Randomizing or preparation of unfamiliar exam questions to each candidate would curtail exam irregularity online. It is not very easy for students not to congregate to brainstorm on or discuss exam questions remotely during online assessments or examinations. Having candidates present online would be an added remedy to seal off all possibilities in online cheating. Proctoring assessment which is handled remotely, deployed by the Learning Management System (L.M.S.) 's Canvas, Sakai, or Proctor track application, would supplement these approaches because they are system timed, providing invigilators with controls in monitoring the candidate remotely during sessions.

#### ***6.1.9 Alternative power sources***

According to the economy of energy in Sub-Saharan Africa, despite the abundance of minerals and natural energy sources, matched with enormous and desperate demands for energy, Sub-Saharan Africa remains significantly un-electrified. An energy mix and utility reform can be the solution, Ramonyai (2019). Innovation attempts are stagnating because of power interruptions. Our local Internet Service Providers (I.S.P.) are similarly experiencing power outages from the main, Kenya Power and Lighting Company (KPLC). Once I.S.P. connectivity

equipment is down, consumer connections come to a grinding halt. Probably, during the Covid-19 pandemic coupled with the current digitalization, Kenya should prepare to develop more permanent alternative energy sources, such as solar pond technology and more so, the nuclear power plants. At nuclear power plants, neutrons collide with uranium atoms, disintegrating them at the core of nuclear reactors. This enables the fission of uranium atoms to release energy that heats waters. The heated waters are used to spin turbines that generate cleaner energy (electricity) to the national grid. Such a wonderful project would sustain energy power supply that would support vital industries like education and healthcare at their critical data centers while maintaining greener cities. The figure below represents a nuclear power plant as a better alternative energy power source, which should be a focus.



**Figure 2:** Nuclear power plant, an alternative power source

#### *6.1.10 Resilient I.T. team*

Locally, I.T. teams remain the ones tasked with the road map for mounting an online education platform. The team requires enough support necessary for the sustainability of online education. In the event that they give up, the whole process would stall. The I.T. team should be resilient enough to jump-start the processes following the impact of resistance from both faculty and students. A part of the non-I.T. faculty members should be constituted as a technical committee to represent a cross-section of employees and constituted as the security design team to work with I.T. to help alleviate change resistance.

#### *6.1.11 Adherence to security design and implementation*

Currently, computerization and rationalization are the picks of desired change in any competitive organization; although they are slow in changing and moving strategies, they are expected to bring about modest returns on investments and little risk in organizations. Using Learning Management Systems (L.M.S.) for online education should be an all-inclusive change strategy towards carrying greater returns on investment for education; although it also comes with high risks and chances of failure, much caution MUST be taken for this venture to be fruitful.

#### *6.1.12 Technical security control measures*

Technological controls exist in the form of access controls and authentications. They establish who has and who does not have access to education information resources and makes sure that users are who they purport to be when accessing information resources; this is the main role for technical controls. However, it is worth noting

that the prevailing trends where users prefer single and simple PASSWORDs are very inadequate. Mechanisms for data breaches require more than one PASSWORD because if one PASSWORD is stolen, the other will remain unknown to the attacker.

Notably, technical security controls are good measures for securing sensitive information. Leaks rarely stem from breaches of technical controls. Entertaining social engineering or finding what has been written on paper is far easier than an intrusion into an organization's physical I.T. infrastructure. Technical controls require management backing with a formal rule-based structure that considers even hiring guidelines and human resources responsibilities. This establishes an infrastructure of responsibility, leading to attributing blame, responsibility, accountability, and authority.

Access control would standardize, impound, observe, and finally offer protection to the education resources online. Layered security that constitutes more security measures for filtering attack incidences and threats, is to assure strategy to secure private from the public network, so that in the event that one defense becomes faulty, the other measures become a fallback guard. Layered security includes firewalls, malware scanners, Intrusion Detection Systems (I.D.S.), integrity auditing procedures, and even local storage encryption tools to protect information resources in a means other single protection strategies will not. A layered security strategy is an important and surest means to protecting online education platforms.

#### **6.1.13 The legal framework**

It is time for Sub-Saharan Africa to be aware of cybercrime challenges and to begin amending its legislation. This will ease the admissibility of cybercrime matters in our courts of law. Kenya and other East African member states are still in limbo on cybercrime war, even though there is potential in Kenya because of the new data protection law in place, already signed into Laws of the Land a year ago, this MUST be supported by legislative instruments, law enforcement agency and jurisdiction/prosecution to evolve further to respond to the emerging cybercrime challenges. It is even of no use if legislation cannot be enforced because culprits will always escape. Relevant structures have to be available for cybercrime law enforcement and mechanisms with enough resources through which enforcement can be operated to make offenders culpable for punishment. This will act as deterrence to potential criminals' conducts online.

Education stakeholders should work closely with cyber security specialists together with legislative bodies to make cyber laws for global education in recognition of cybercrime challenges and to amend legislation to affect them.

### **Acknowledgements**

It is my pleasure to acknowledge first of all Our Creator Almighty God in Heaven, His Angels and all Saints, in a special way St. Isidore, the patron saint of computers and their users, programmers, repair people, as well as the Internet as a whole, for the knowledge and strength that I use all along. The efforts, support, and encouragement from my wife Elizabeth, my children Gladwell, Gloriachrista, Gregory & Gerry-Jerome, my dad Pius Omoga and especially my late mother Calsine Aketch Omoga for having taken good care of me and her ever greater vision towards my education and all members of my family. Founding Chancellor of Uzima University-Most Rev. Zaccheus Okoth Arch Bishop Emeritus of the Catholic Metropolitan Sea of Kisumu. Rev. Dr. Cosmas Rhagot K'Otieno VC Uzima University. My Research Mentors/Supervisors Prof. Antony J. Rodrigues and Dr. Silvanice O. Abeka (JOOUST). My lecturers Dr. David Halfpenny, Charles Ayoma-McGrath-Australia, Dr. Ogara Solomon, Adv. Hannes Bezuidenhout-University of Pretoria Z.A., for their tireless and high-quality skills and guidance during my research and studies towards the realization of this paper. More encouragement and support from work colleagues Maurice Onywera Medical Lab technologist, Dr. Alfred Osoro – School of Nursing. Dennis Wanda and Kevin Otieno – I.T team at the I.S/I.T Dept. at the Uzima University and all my medical students, especially Medical Informatics" Bachelor of Medicine & Bachelor of Surgery (MBChB) 2019 and 2020 classes at The Uzima University.

### **References**

- Abrahamson, S. D., Canzian, S. and Brunet, F., 2005. Using simulation for training and to change protocol during the outbreak of severe acute respiratory syndrome. *Critical Care*, 10(1), Pp.1- 6.
- Axelrod, C. W., 2015. Enforcing security, safety and privacy for the Internet of Things. In: *IEEE, 2015 Long Island systems, applications and technology*.

- Bradley Morrison, J. and Rudolph, J. W., 2011. Learning from accident and error: avoiding the Hazards of workload, stress, and routine interruptions in the emergency department. *Academic Emergency Medicine*, (12), pp.1246-1254.
- Bretag, T. 2016. Discipline-specific approaches to academic integrity: Introduction. *Handbook of Academic Integrity*, 673-675.
- Bretag, T. ed. 2016. *Handbook of academic integrity*. Singapore: Springer.
- Brown, M., Haughwout, A., Lee, D., Scally, J. and Van DerKlaauw, W, 2015. Measuring student debt and its performance. *Student loans and the dynamics of debt*, pp.37- 52.
- Chen, S. and Macfarlane, B., 2016. Academic integrity in China. *Handbook of academic integrity*. Singapore: Springer. pp.99-105.
- Devi, S., 2020. Economic crisis hits Lebanese health care. *The Lancet*, 395(10224), p.548. Eaton, S. E. 2020. Academic integrity during COVID-19: Reflections from the University of Calgary.
- Eppich, W., and Cheng, A., 2015. Promoting Excellence and Reflective Learning in Simulation (PEARLS): development and rationale for a blended approach to health care simulation debriefing. *Simulation in Healthcare*, 10(2), pp.106-115.
- Fishman, T., 2016. Academic integrity as an educational concept, concern, and movement in US institutions of higher learning. *Handbook of academic integrity* Singapore: Springer. pp.7-21.
- Foster, G., 2016. Grading standards in higher education: Trends, context, and prognosis. *Handbook of academic integrity*. Singapore: Springer. Pp.307-324.
- Gamage, K. A., Silva, E. K. D., and Gunawardhana, N., 2020. Online delivery and assessment during COVID-19: Safeguarding academic integrity. *Education Sciences*, 10(11), p.301.
- Gilmore, J., Maher, M. and Felton D., 2016. Prevalence, prevention, and pedagogical techniques: Academic integrity and ethical professional practice among STEM students. 2016. *Handbook of academic integrity*. Singapore: Springer, pp.729-748.
- Gilmorea, J., Mmaherb, M. and Fedonc, D., 2015. Prevalence.Q1 Prevention and Pedagogical Techniques: Academic Integrity and Ethical Professional Practice among STEM students. *Handbook of academic integrity*, 729-748.
- Greenberg, H. J., and Maybee, J. S., Eds...2014. *Computer-assisted analysis and model simplification: Proceedings of the first symposium on computer-assisted analysis and model simplification, University of Colorado., Boulder, Colorado, 28 March, 1980*.Elsevier.
- Huang, Y., Huang, Z., Zhao, H., and Lai, X., 2013. Anew one-time password method. *IERIProcedia*, 4, pp.32-37.
- Huang, Y., jiaXue, W., shi Huang, G. and jia Lai, X., 2013. On the security of multi- factor authentication: several instructive examples. In: *Proceedings of the 2013 International Conference on advanced Computer Science and Electronics Information*. Atlantis Press.
- Kim, H., Krishnan, C., Law, J and Rounsaville, T., 2020. *COVID-19 and US higher education enrollment: Preparing leaders for fall*. New Jersey: McKinsey &Company.
- Lamport, L. 1981. Password authentication with insecure communication. *Communications of the ACM*, 24(11), pp.770-772.
- Morris, E.J., and Carroll, J. 2016. Developing a sustainable holistic institutional approach: Dealing with realities 'on the ground' when implementing an academic integrity policy. *Handbook of academic integrity*. Singapore: Springer. pp.449-462.
- M'Raihi, D., Bellare, M., Hoornaert, F., Naccache, D. and, Ranen, O., 2005. Hotp: An HMAC based one-time password algorithm. RFC4226. The Internet Society, Network working Group.
- Mumtaz, G., 2020. Providing context for COVID-19 Numbers in the Arab region. *Nat Middle East*, p.10.
- Ning, H., Liu, H. and Yang, L. T., 2013. Cyber security in the Internet of things. *Computer*4646 (4), pp.46-53.
- Nwuke, T. J. And Gberepikima, J.E., 2020. The impact of COVID-19 on the Educational system in Nigeria *People*, 3(3), pp.51-68.
- Omogah, F. O., 2020. The embryonic COVID-19 themed cyber threats, a looming tragedy to already vulnerable Global Electronic Healthcare Systems (GEHCS). *Health Systems and Policy Research*.7 (5), pp.2254-9137.
- Orim, S. M., 2016. Perspectives of academic integrity from Nigeria. *Handbook of academic of integrity*. pp.147-169.
- Pecorari, D., 2016. Plagiarism, international students and the second-language writer. *Handbook of academic integrity*. Singapore: Springer. pp.1-11.
- Ramonyai, M., 2019. The economy of energy in Sub-Saharan Africa. *Business Unusual Quarterly Energy Edition Journal*: pp.8-9.
- Raj, R, Wong, S. H. and Beaumont, A. J., 2016. Business intelligence solution for SME: A case study.
- Saad, M. and Soomro, T. R. 2018. Cyber security and the Internet of things. *Pakistan journal of Engineering, Technology*.
- Schwartz, J., King, C. C., and Yen, M.Y., 2020. Protecting healthcare workers during the Corona virus disease 2019 (COVID-19) outbreak: Lessons from Taiwan's severe acute respiratory syndrome response. *Clinical Infections Disease*, 71(15), pp.858-860.
- Slay J. and, Koronios, A., 2006. *Information technology security and risk management*. Willey.
- Stephens, J. M., 2016. Creating cultures of integrity: A multi-level intervention model for promoting academic honesty. *Handbook of academic integrity*. Singapore: Springer. pp.996-1007.
- Sou, H., Liu, Z., Wan, J. and, Zhou, K., 2013. Security and privacy in the mobile cloud computing. *9<sup>th</sup> International wireless communications and mobile computing conference*.
- Sou, H., Wan, J., Zou, C., and, Liu, J., 2012. Security in the Internet of things: a review. In: *2012 international conference on computer science and electronics engineering* (Vol. 3, pp. 648-651).



**Fredrick Ochieng' Omogah**

- Weber, R.H., and Studer, E., 2016. Cyber security in the Internet of things: Legal aspects. *Computer Law & Security Review*, 32(5), pp.715-728. World Health Organization. (2020). *Responding to community spread of COVID-19: interim guidance*, 7 March 2020 (No. WHO/COVID-19/Community Transmission/2020.1). World Health Organization.
- Wright, G. H., (1942). Management of information security Chapter 7 risk Management: Identifying and assessing Risk  
Once we know our weaknesses, they cease to do us any Harm
- Zeiler, K., 2016. The future of empirical legal scholarship: Where might we go from here? *J. Legal Educ.*, 66, p.78.
- Zhang, Z., Cho, M., Wu, Z., and Shieh, S. W., 2015. Identifying and authenticating IoT objects in a natural context. *IEEE annals of History of Computing*, 48(08), pp.81-83.

# How Penetration Testers View Themselves: A Qualitative Study

Olav Opedal

Opedal Consulting LLC, Ellensburg, USA

[olav@opedalconsulting.com](mailto:olav@opedalconsulting.com)

DOI: 10.34190/EWS.21.058

**Abstract:** Many organizations understand they need penetration testers to identify network security weaknesses. In response, penetration testing has become required for most major organizations. As a result, penetration testing became a cyber security occupation. Penetration testers perform activities similar to criminal hackers with one major difference: penetration testers work on behalf of the organization and they do not attempt to criminally exploit the organization. Earlier research identified that penetration tester personality traits are different from other computer professionals personality traits. Little is known about how professional penetration testers view themselves. Little research exists studying the use of hacking methods from the perspective of the penetration testers themselves. This qualitative study sought to increase understanding about penetration tester traits by exploring the views of penetration testers. The study included interviews with thirteen red team members at a global software firm in the Pacific Northwest.

**Keywords:** qualitative content analysis, latent Dirichlet allocation, topic modeling, penetration testers

---

## 1. Introduction

Information security is a major concern for most organizations. Most companies have transformed to a digital environment or they are in the process of undertaking the transformation to a digital environment. Furthermore, as part of the transformation, organizations must adapt defense methods as they now face organized and disciplined threat actors (Tounsi & Rais, 2018). Criminals who attack computer systems for notoriety or financial gain are referred to as *hackers*. However, not everyone referred to as a hacker is a criminal. Those hackers who do not commit crimes are referred to as ethical hackers (Sinha & Arora, 2020). Penetration testers are white hat, ethical hackers, and they are hired by organizations to identify vulnerabilities in computer systems (Hatfield, 2019; Sinha & Arora, 2020). Black hat hackers are the criminal hackers, and gray hat hackers fall in-between (Hatfield, 2019). Crosston (2017) argues that cyber disobedience should be separated from other cyber crime. The goal of hackers is to obtain access to computer systems without being detected by the organization's defenses (Sinha & Arora, 2020). Hackers can use technical methods or using psychological techniques to obtain information from human victims (Sinha & Arora, 2020). The information enable a hacker to access computer systems (Hatfield, 2019). The use of psychological techniques to obtain information using a ruse is called social engineering (Hatfield, 2019). Penetration testers can use technical methods, social engineering methods, or a blended approach to meet their objectives. Social engineering is the practice of manipulating human victims to gain access to privileged information or direct access to computer systems or networks (Hatfield 2019). Furthermore, it is common to distinguish between white hat and black hat hackers, which is determined by the motivation for the hacking activity and if the victim provided prior consent (Hatfield 2019). The motive for researching the psychology of hackers is that psychological research into hackers enable increased understanding of motivations driving the hacker (Thackray et al. 2016).

## 2. Rationale for the study

The rationale for studying the psychology of professional penetration is the recognition that penetration testing is an emerging occupation. The emergence of penetration testing as a profession is evidenced by the inclusion of the occupation by the United States Department of Labor into their O\*NET data base of occupations (National Center for O\*NET Development, 2020). The work conducted by penetration testers is recognized by the United States Department of Labor (DOL) as an occupation with a bright outlook (National Center for \*ONET Development, 2020). The DOL noted that the median yearly wage for a penetration tester in the United States is \$88,550 (National Center for \*ONET Development, 2020). Tasks commonly performed by penetration testers include physical security assessments, the development of risk mitigation strategies, the performance of security audits, implementation of least privilege access to computer systems, ensuring that only essential system features have been enabled, and creating solutions that mitigate known vulnerabilities (National Center for \*ONET Development, 2020). Yagoob et al. (2017) noted that there are two types of penetration testing. The first type is physical penetration testing where a penetration tester tests the physical security of physical assets such as data centers, offices, and network equipment. The second type of penetration testing occurs over the network. Adamović, Božić, and Penevski (2019) described penetration testing as a complete set of

methodologies used to assess the security posture of an organization. Adamović et al. (2019) referred to penetration testers as ethical hackers. Adamović et al. (2019) noted that the methodology is broadly grouped into five phases: (a) target recognizance; (b) a scan of the publicly facing network systems, applications, and databases; (c) evaluation of exposed vulnerabilities, (d) exploitation of vulnerabilities, and (e) provision of a report of the findings from the vulnerability assessment. Bertoglio and Zorzo (2017) conducted a meta study of 54 primary studies of penetration testers to classify the tools and models used for penetration testing. Bertoglio and Zorzo (2017) noted that penetration testing is conducted in three phases, a) pre-attack, b) gaining entry, and c) the post attack phase. Furthermore, there are three approaches to penetration testing, a) a manual exploratory approach, b) an automated approach, and c) a systematic manual approach (Bertoglio & Zorzo, 2017). Yan (2020) noted that a key question for the scientific community is how humans interact with cyber technologies. The goal of the research of professional penetration testers was to elucidate what hacking means to professional penetration testers by capturing professional penetration testers' perspectives and experiences. This study sought to answer a small part of the broader research question Yan (2020) proposed by better understanding how penetration testers viewed hacking and their own role in hacking.

### **3. Hacker personalities**

Turgeman-Goldschmidt (2008) explored hackers' narrative stories and found that they generally fall into ethical versus unethical hacker categories. Turgeman (2011) noted that the term *hacker* is ambiguous and can mean computer experts, or it can mean those who break into computer systems, or create pirated software copies, or those who steal credit card information. Researchers use self-control theory to understand the difference between white hat hackers and black hat hackers. According to Marcum and Higgins (2014) self-control theory provides support for understanding cyber-crime, and unethical hackers lack the self-control exercised by ethical hackers. Marcum and Higgins (2014) noted that motivations for hacking included addiction, curiosity, excitement, entertainment, money, power, status, ego, ideologies, peer recognition, and revenge. According to Dehaene (2020), curiosity is a key driver for the motivation to learn. Curiosity and motivation to learn are associated with the brain's reward centre (Dehaene, 2020). Marcum and Higgins (2014) noted that self-control is a stable trait over time, and self-control is a good predictor of criminal behavior, both online and off. Ethical hackers with self-control are more likely to become professional ethical hackers. In studies about personality, researchers found that penetration testers were more honest and humbler compared to other computer professionals (Opedal, 2019). Gaia et al. (2020) found that the difference between the personality traits of white hat, grey hat, and black hat hackers and found that the dark triad of personality traits, narcissism, psychopathy, and Machiavellianism was a predictor of a person being a hacker, but not the type. Gaia et al. (2020) found that opposition to authority was a predictor that could identify gray hat hackers when compared to white hat and black hat hackers. Originally, researchers developed five personality trait dimensions: extroversion, agreeableness, conscientiousness, openness, and neuroticism (McCrae & Costa, 2003), but the dimensions were later extended to include honesty and humility in the Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness, HEXACO, model of personality (Ashton & Lee, 2020). The HEXACO model share extraversion, conscientiousness, and openness, with the Big 5 model but considers neuroticism and agreeableness to be better explained as the following three dimensions: honesty-humility, emotionality, and agreeableness versus anger compared to the original neuroticism and agreeableness (Ashton & Lee, 2020). Another important aspect of hacking is reputation. According to Przepiorka, Norbutas, and Corten (2017), an online reputation with underground drug markets was correlated with trust; therefore, those with higher levels of trust could demand higher prices. Gambetta (2011) noted that criminals use signaling to convey trust to conduct business in more traditional criminal marketplaces. Understanding the personality, values, and behaviors of hackers can help build a typology for these individuals. Typology, the creation of a categories to enable classification of personality types, parenting types, and coping strategies have a long tradition in psychology (Stapley, O'Keefe, & Midgley, 2021). Seebruck (2015) argued that a typology of hackers can be used to represent the multifaceted composition of hacker types. Marin, Shakarian, and Shakarian (2018) argue that most hackers are unskilled and uninteresting from a research perspective, but those that are at the top should be investigated. However, Bruijne, Eeten, Gañán, and Pieters (2017) lament that there is a lack of a commonly agreed upon typology of hackers, but identified from their search of the literature that the following five dimensions exists: a) target, b) expertise, c) resources, d) organization, and e) motivation.

### **4. Hacker sub culture**

Thackray et al. (2016) argued that social psychology is an underutilized tool in cyber security, and psychological research into hackers enable increased understanding of motivations driving the hacker (Thackray et al. 2016).

Thackray et al. (2016) notes that white hat hackers tend to be motivated by prestige and that curiosity and thrill seeking are motivators for hackers in general. Turgeman (2011) argued that the term *hacker* is a socially constructed term and that hackers view themselves as positive deviants. Nycyk (2016) found that those who want to learn hacking skills face significant challenges within the hacker community. Participants in the hacking community deliberately block others from learning hacking techniques (Nycyk, 2016). The obstacles hackers use to block others from learning hacking increase the cost of learning hacking methods and making entry into the subculture more difficult. Berger (2016) argued that identity signaling comes with a cost, such as invested time (Berger, 2016). If those in the know create hurdles for others trying to gain the knowledge necessary to attack computer systems, they significantly increase the interested individual's time it will take to learn the trade. Anyone can claim to be a hacker or claim to be knowledgeable in the skills associated with the hacking, and the demonstration of knowledge can set hackers apart from general computer users. Gambetta (2011) noted the importance for criminals to signal others that they are trustworthy criminals and they do so by making the cost of entry into the criminal subculture expensive. The costs of entry into the subculture of hacking are also expensive, and hackers are viewed by society as deviants (Turgeman, 2011). Choosing to join a group (e.g., hackers), and the behavior subsequently displayed as part of a group, are based on socially normative factors (Matias, 2019). Matias (2019) conducted a large-scale random experiment where participants were given announcements for acceptable behavior. Matias (2019) found that behavior change depending on the messaging they saw when they joined online social networks. Berger (2016) noted, for example, that in societies where bottle-feeding children was associated with a mother's HIV infection, infected mothers chose to breastfeed their children in public instead of bottle-feeding them due to the stigma associated with HIV. Maxigas (2017) noted that self-described hackers would go to significant lengths to live up to their commitments to only use certain technologies despite the extensive time it took to perform the behavior. Rajadesingan, Resnick, and Budak (2020) noted that self-selection, having the necessary prior knowledge, and the willingness to continue to learn new knowledge while a member of an online social network were key components of sustained membership. Rajadesingan et al. (2020) further argued that communities with norms differing from more mainstream norms strive to ensure that those who are compatible with the in-group members are attracted to and retained within the group. Those who want to become a successful penetration tester must learn how to exploit security vulnerabilities in computer systems, in networks, or via social engineering. Learning how to use the tools of a hacker requires self-direction and motivation, especially in an adverse environment (Nycyk, 2016). Seo and Patall (2020) found that those with high levels of self-control study even harder when they experienced negative emotions while trying to learn new skills.

## **5. Summary**

The review of the literature about online communities, and about hacker communities more specifically, supports the idea that hacking is also a sub-culture. The hacking sub-culture values curiosity and a willingness to learn how to use technology in unanticipated ways contrary to intended and socially accepted uses. The hacking sub culture includes those who self-select into hacker communities despite, or perhaps because of, the high bar associated with entry. The review of the literature also identified that hackers tend to fall into white hat vs. black hat hackers, with a smaller population of grey hat hackers.

## **6. Method**

Data were gathered via interviews with thirteen penetration testers employed at a large, multinational software company based in the Pacific Northwest. Participants were identified using a snowball method where each interviewee was asked which other penetration testers they thought should be interviewed. The sampling approach followed the traditional snowball sampling method outlined by Handcock and Gile (2011). The objective of this study was to explore how penetration testers described their lived experiences with, and interpretations of, hacking. The study used the grounded theory (GT) method formulated by Strauss (1987) and Strauss and Corbin (1990, 2000). Strauss and Corbin (1990, 2000) noted that researchers that use GT do so to derive theory, and the resulting text is the interpreted reality. Ryan and Bernard (2003) described how theme identification is the cornerstone of qualitative research. Elliott and Timulak (2021) noted that qualitative research should be referred to as generic descriptive-interpretative qualitative research (GDI-QR). Qualitative research methodology includes letting the answers to open ended questions guide the study, to collect the experiences and observations of the subjects, complete a full analysis of descriptions and observations, that meaning is represented through the analyst's understanding of the subjects' experiences, clustering themes, disclosure of theory, and finally, creating a coherent model or story (Elliott & Timulak, 2021).

Latent Dirichlet allocation (LDA), was first proposed by Blei, Ng, and Jordan (2003). Blei (2012) described how LDA is a method used to investigate scientific text with Bayesian learning methods. Blei argued that LDA could be used to discover themes that could separate text from different scientific disciplines through the themes identified in texts for a specific scientific subject. Blei (2012) noted that a document often consists of a mixture of topics. Furthermore, Blei (2012) chose to use a mixed membership model rather than a hard cluster model, which allowed a word to belong to more than one cluster to preserve the betweenness in between words.

Baumer et al. (2017) compared LDA to GT and found that LDA performs well in comparison with the traditional GT approach. This study differed from the GDI-QR process in that clustering was accomplished using the machine learning method (i.e., LDA) as proposed by Baumer et al. (2017). Baumer et al. (2017) noted that output from the LDA model must still be interpreted by the researcher. Blei, Ng, and Jordan (2003) found that the basic method used to extract information from documents required that the documents be converted into a corpus of vector ratios of word counts. Latent Dirichlet Allocation algorithm relies on a theorem developed by de Finetti (1990) where exchangeable random variables can be represented as a mixture distribution (Blei et al., 2003). Blei et al. (2003) applied de Finetti's theorem to identify the statistical structure between the documents in a corpus. Blei et al. (2003) found that using the mixing distribution resulted in the capture of topics contained within the corpus. Blei et al. (2003) noted that each document is repeatedly sampled for each topic node resulting in the possibility of a document being linked with multiple topics. Baumer et al. (2017) compared and contrasted the interpretative method from social sciences with the LDA model from the field of machine learning. Bauer et al. (2017) argued that machine learning methods work beyond human capabilities in the size of text that can be analyzed, but a limitation is that the statistically defined topics may be misleading to a human. Understanding the context of the social reality of the subjects of interest can be lost (Baumer et al., 2017). Mitchel (2019) explained that the current state of the art natural language processing (NLP) does not have the ability to comprehend words and sentences due to the lack of a mental model necessary to understand text in the context of a lived experience.

Earlier psychological studies used topic modeling to determine if there were meaningful differences between cultural subgroups. One example was a study conducted by Sundararajan, Ting, Hsieh, and Kim (2020) that used topic modeling to evaluate cultural differences between two religious' groups: Christians and the Bimo religion of the Yi ethnic minority. Sundararajan et al. (2020) argued that machine assisted analysis provided advantages over traditional qualitative methods after comparing the results of the text analysis from machine learning with the results from manual coding. Dehaene (2020) noted that caution is advisable when using machine learning techniques to analyze text. Dehaene (2020) stated that machine learning cannot determine the meaning of the text, and human interpretation of the results is necessary. Machines do not comprehend the gestalt expressed in text.

## **7. Participants**

The participants were employed as penetration testers at a large software company located in the Pacific Northwest. The participants worked for the company at the time of the interviews in one of the read teams. The participants were interviewed at their workplace using teleconferencing software. Each participant was asked for permission to record the interview and was promised anonymity. Each interview was transcribed by a professional transcriber. The interviews lasted an average of one hour. The participants were all males ranging in age from their early 20s to their mid 40s at the time of the interviews. No female participants were available for interviews. All but one of the penetration testers self-identified as a hacker. Twelve were Caucasian and one was of Asian descent. Two of the participants held dual roles as managers of their respective teams while also working as penetration testers.

## **8. Results**

Each interview began with the request for approval to record the conversation. The transcribed interview data were loaded as text files into a Python pandas data frame. The participants' names were replaced with numbers when the data were loaded into the data frame. The next part of pre-processing was to normalize the text, which meant that all the words were converted to lowercase. Normalization was followed by removing common English stop words, removing the punctuation, and finally lemmatizing the words. The process resulted in a collection of 3997 unique tokens derived from the corpus. The final step was to create thirteen individual document term matrices that were loaded into an LDA model, and the model was then generated. Topics are represented as word distributions (Blei, 2012). Topic models derived from LDA can be difficult to interpret

without visual aids that provide interactivity with the fitted LDA model (Sievert & Shirley, 2014). Interactive visualizations of the fitted LDA model can help the researcher identify the meaning of each topic, the prevalence of each identified topic, and finally, how the topics were related (Sievert & Shirley, 2014). The LDA analysis identified three distinct topics: a) hackers think differently, (b) hacker culture is different, and (c) hackers use curiosity and imagination to explore boundaries. The main theme is that hackers differ from others in thinking, culture, and behavior. The sub-topics were a) hackers must think like a hacker, not like other people, (b) the hacking sub-culture consists of motivated individuals attempting to understand systems, and (c) hackers pursue knowledge, sometimes to the extreme, to obtain control, to seek new experiences, and to explore technological boundaries.

## **9. Evidence that hackers think differently**

Subject 1 described how hackers think by saying, "You're kind of willing to just think in a different way with just like a natural inborn curiosity I guess is probably what I would think a hacker just kind of has by default." Subject 5 described introverts thusly, "I guess from my experience I'd probably put most of us in the introvert side." Subject 1 said "I'm not the only person who walks into a room at some reception or get together and is like this is like pulling teeth, I hate this, I don't want to be here, but you do it because you have to." Subject 5 described thinking differently by saying, "I think that one, you know, especially as you're in a role like this or in this industry and focusing in this area, you tend to realize that there are things that the vast majority of people don't bother to question either because they don't think they have anything to fear or it's just not part of their daily experience or again because there's that kind of definition of a hacker, at least mine going back to being intellectually curious and kind of questioning things as you go, you know, I think a lot of people in my role, for example, would look at a privacy statement and wonder if they're checking the customer permit box what data is being sent and what who would have access to that and what could they glean from that. And then possibly even going further and saying how can I get access to that. Whereas a typical user doesn't bother to uncheck the checkbox."

## **10. Hacker culture**

Subject 1 described how he felt about the hacker culture when he said, "I think there's definitely a culture to hackers," and Subject 1 felt the hacker culture was similar to addiction when he said, "I mean if I don't have my phone on me, I feel like something's missing, and that's kind of sad, I don't like it, but I definitely do feel like I have somewhat of an addiction to it. And I definitely know other people fall into the same category." Subject 4 said, "So people who I would label hackers, a large portion of them actually never finished college, never fit well into institutionalized study. Never very well went along with what people expected of them and very hand in hand with that type of scenario, they're very strong out of the box thinkers."

## **11. Exploring Boundaries**

Subject 1 described exploring the boundaries of technologies when he said, "What makes up a hacker I think it's somebody that has an interest in figuring out what a system is capable of doing, not necessarily what is intended to do." Subject 2 talked about computer system knowledge by saying, "I believe a hacker essentially is someone who within the context of a computer has a very deep understanding of some system which they may be or software that they may be trying to use to the point where they're able, where they know the insides and outs of it so that they can potentially make it do something that it initially wasn't intended to do, whether this is for good or nefarious purposes I guess." Subject 3 talked about gratification as motivation when he said, "But there are factors that are much more deep than that, there's definitely a sense, a very strong sense of gratification. And gratification of being able to accomplish what you set out to do because breaking into things definitely is first and foremost a challenge, then once you actually get in it's an accomplishment." Subject 5 said, "Certainly there's probably some concern that you're doing something even when you're on the white hat side that you might do something that crosses a boundary that you weren't supposed to."

## **12. Discussion**

The themes that emerged from the analysis of the interviews of penetration testers are supported by the results of earlier studies of hackers. The current study differed from other research in that those studied were all working as penetration testers at a large US software company. The study found that white hat hackers also see themselves differently from others. Turgeman (2011) noted that, in general, hackers view themselves as different from others. This study confirmed that penetration testers see themselves as hackers and different from others even though they worked in a professional capacity and not as criminals. The study further found

that penetration testers view hacking as a separate sub-culture, and they identified themselves as part of that sub-culture. Nycyk (2016) found that knowledge seeking was important to hackers. Knowledge seeking was also found as a theme in the study of penetration testers. Maxigas (2017) described hackers' motivations to explore the boundaries of technology, but the results identified that the penetration testers in this study were also motivated to explore boundaries. The causes for the difference in thinking about systems between hackers and the general population may be found in the composition of these individuals and their sub-culture; however, penetration testers tend to be more introverted than most other computer professionals (Opedal, 2019). Evidence exists that the process of self-selection into hacking sub-cultures is the same for ethical hackers as for criminal hackers. According to Marcum and Higgins (2014), self-control predicts who is likely to become a criminal hacker. The study of the personality of penetration testers found that penetration testers are more likely to be honest and humble when compared to other computer professionals (Opedal, 2019); therefore, high scores on the honesty-humility facet of personality is likely a key factor for who becomes a criminal hacker and who ends up employed as a penetration tester. Thackray et al. (2016) noted that paranoia is common among hackers. The data from the interviews also identified paranoia as a concern among some of the interviewees.

### **13. Implications of findings**

This study of penetration testers has several managerial implications. The results indicated that penetration testers follow the same processes to become hackers that grey hat hackers and black hat hackers follow. The study added to the understanding of hacker typologies. Prediction of group membership becomes possible when a typology can be assigned to common behaviors that differ between different groups of individuals (Stapley, O'Keefe, & Midgley, 2021). According to the United States Department of Labor (2020), the need for penetration testers is expected to rise in the United States. Because penetration testers self-select into the profession, it is important to be able to measure their skills and aptitudes prior to hiring. It is also important to measure their scores in the honesty-humility dimension. Furthermore, introversion is a key trait found among penetration testers, and penetration testers may need more accommodation to function in corporate environments. Measuring personality traits, skills, and aptitudes should enable management to make informed hiring decisions and eventually become better secured against internal and external attacks. It is important to note that if the wrong person is hired to be a penetration tester, that person can cause significant harm to the organization. An insider attack from someone highly skilled and motivated is likely to stay undetected for a longer period of time, and the hacking is likely to cause significant harm as the insider has in-depth knowledge about the organization. The main difference between ethical and criminal hackers was the level of self-control (Marcum & Higgins, 2014), and self-control can be measured with a personality inventory (Opedal, 2019).

### **14. Evaluation**

The study used a convenience sample of individuals that at the time worked as penetration testers at a very large global software company. These individuals were all male, and mostly Caucasian. They were all successful professionals, some early career, others mid-career. Due to the limited sample size and that they were all employed by the same US based software company limits the generalizability of the findings. Another important limitation to the study is that the output from the LDA model had to be interpreted. It is possible that the interpretation and subsequent labelling of the resulting word distribution is flawed. The findings were similar to prior studies, and therefore, one can draw the conclusion that LDA proved to be an effective method for the study of penetration testers.

### **15. Recommendation for future research**

Future research of penetration testers should explore their motivations and personalities in more depth. Earlier research into the personality traits of penetration testers used the short personality instrument, the miniIPIP6 (Opedal, 2019). Using other instruments could support or refute the results in this study. Furthermore, no studies to date have explored the life experiences of female penetration testers or if there are gender differences in personality traits among hackers.

### **16. Conclusion**

Research exists into who hackers are, but knowledge of those employed as penetration testers is limited. Earlier research provided important insights into the skills they need to obtain, how they obtain the necessary skills, and a theoretical foundation for why criminal hackers commit crime. A theory of white hat hackers, however, was not found. The thematic analysis found that hackers, including penetration testers, differ from others in thinking, culture, and behavior. All but one of the penetration testers studied self-identified as hackers. The

three topics that were identified were (a) penetration testers must think like hackers, (b) penetration testers are, for the most part, members of the hacking sub-culture, and (c) the hacking sub-culture consists of highly motivated individuals whose main goal is to understand computer systems. Finally, hackers pursue knowledge, sometimes to the extreme, to obtain control, to seek new experiences, and to explore technological boundaries. What separates ethical from criminal hackers is self-control (Marcum & Higgins, 2014).

## References

- Adamović, S., Božić, K. & Penevski, N., 2019. Penetration Testing and Vulnerability Assessment: Introduction, Phases, Tools and Methods. In *Sinteza 2019-International Scientific Conference on Information Technology and Data Related Research* (pp. 229-234). Singidunum University.
- Ashton, M.C. and Lee, K., 2020. Objections to the HEXACO model of personality structure—And why those objections fail. *European Journal of Personality*, 34(4), pp.492-510.
- Baumer, E.P., Mimno, D., Guha, S., Quan, E. and Gay, G.K., 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence?. *Journal of the Association for Information Science and Technology*, 68(6), pp.1397-1410.
- Berger, J., 2016. *Invisible influence: The hidden forces that shape behavior*. Simon and Schuster.
- Bertoglio, D.D. and Zorzo, A.F., 2017. Overview and open issues on penetration test. *Journal of the Brazilian Computer Society*, 23(1), pp.1-16.
- Blei, D.M., 2012. Probabilistic topic models. *Communications of the ACM*, 55(4), pp.77-84.
- Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3, pp.993-1022.
- Bruijine, M.D., Eeten, M.V., Gañán, C.H. and Pieters, W., 2017. Towards a new cyber threat actor typology.
- Crosston, M., D., (2017) *The Fight for Cyber Thoreau: Distinguishing Virtual Disobedience from Digital Destruction*. Korstanje, M.E. ed., 2016. Threat mitigation and detection of cyber warfare and terrorism activities. IGI Global.
- Dehaene, S., 2020. *How We Learn: Why Brains Learn Better Than Any Machine... for Now*. Penguin.
- Elliott, R. and Timulak, L., 2021. *Essentials of Descriptive-Interpretive Qualitative Research*. American Psychological Association. Washington DC.
- Gaia, J., Ramamurthy, B., Sanders, G., Sanders, S., Upadhyaya, S., Wang, X. and Yoo, C., 2020, January. Psychological Profiling of Hacking Potential. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.
- Gambetta, D., 2011. *Codes of the underworld: How criminals communicate*. Princeton University Press.
- Handcock, M.S. and Gile, K.J., 2011. Comment: On the concept of snowball sampling. *Sociological Methodology*, 41(1), pp.367-371.
- Hatfield, J.M., 2019. Virtuous human hacking: The ethics of social engineering in penetration-testing. *computers & security*, 83, pp.354-366.
- Marcum, C. D., & Higgins, G. E. (Eds.). (2014). *Social Networking as a Criminal Enterprise*. CRC Press.
- Matias, J.N., 2019. Preventing harassment and increasing group participation through social norms in 2,190 online science discussions. *Proceedings of the National Academy of Sciences*, 116(20), pp.9785-9789.
- McCrae, R.R. and Costa, P.T., 2003. *Personality in adulthood: A five-factor theory perspective*. Guilford Press.
- Mitchell, M., 2019. *Artificial intelligence: A guide for thinking humans*. Penguin UK.
- National Center for O\*NET Development. O\*NET OnLine Help: Find Occupations. *O\*NET OnLine*. Retrieved January 5, 2021, from [https://www.onetonline.org/help/online/find\\_occ](https://www.onetonline.org/help/online/find_occ)
- Nycyk, M., 2016. The New Computer Hacker's Quest and Contest with the Experienced Hackers: A Qualitative Study applying Pierre Bourdieu's Field Theory. *International Journal of Cyber Criminology*, 10(2).
- Opedal, O., 2019. *Comparing Personality Traits between Penetration Tester, Information Security, and IT Professionals from Two Cohorts* (Doctoral dissertation, Capella University).
- Patton. M. Q. (2015). Qualitative research and evaluation methods. *California EU: Sage Publications Inc*.
- Przepiorka, W., Norbutas, L. and Corten, R., 2017. Order without law: Reputation promotes cooperation in a cryptomarket for illegal drugs. *European Sociological Review*, 33(6), pp.752-764.
- Rajadesingan, A., Resnick, P. and Budak, C., 2020, May. Quick, community-specific learning: How distinctive toxicity norms are maintained in political subreddits. In *Proceedings of the International AAAI Conference on Web and Social Media*(Vol. 14, pp. 557-568).
- Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field methods*, 15(1), 85-109.
- Seebrock, R., 2015. A typology of hackers: Classifying cyber malfeasance using a weighted arc circumplex model. *Digital Investigation*, 14, pp.36-45.
- Seo, E., & Patall, E. A. (2020). Feeling proud today may lead people to coast tomorrow: Daily intraindividual associations between emotion and effort in academic goal striving. *Emotion*. Advance online publication. <https://doi.org/10.1037/emo0000752>
- Shappie, A. T., Dawson, C. A., & Debb, S. M. (2020). Personality as a predictor of cybersecurity behavior. *Psychology of Popular Media*, 9(4), 475-480. <https://doi.org/10.1037/ppm0000247>
- Sinha, S. and Arora, D., 2020. Ethical Hacking: The Story of a White Hat Hacker. *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)*, ISSN, pp.2347-5552.



### **Olav Opedal**

- Stapley, E., O'Keefe, S. & Midgley, N., 2021. *Ideal-Type Analysis A Qualitative Approach to Constructing Typologies*. American Psychological Association, Washington DC. ISBN: 978-1-4338-3453-0
- Sundararajan, L., Ting, R. S.-K., Hsieh, S.-K., & Kim, S.-H. (2020). Religion, cognition, and emotion: What can automated text analysis tell us about culture? *The Humanistic Psychologist*. Advance online publication. <https://doi.org/10.1037/hum0000201>
- Thackray, H., McAlaney, J., Dogan, H., Taylor, J. and Richardson, C., 2016. Social psychology: An under-used tool in cybersecurity.
- Turgeman-Goldschmidt, O., 2011. Identity construction among hackers. *Cyber criminology: Exploring internet crimes and criminal behavior*, pp.31-51.
- Turky, M. and Soliman, N., 2020. Developing auto-Didactic (Self-Learning) Skills by Using Social Networking. *International Journal of Instructional Technology and Educational Studies*, 1(1), pp.16-19.
- Yan, Z., 2020. A basic model of human behavior with technologies. *Human Behavior and Emerging Technologies*, 2(4), pp.410-415.
- Yaqoob, I., Hussain, S.A., Mamoon, S., Naseer, N., Akram, J. and ur Rehman, A., 2017. Penetration testing and vulnerability assessment. *Journal of Network Communications and Emerging Technologies (JNCET)* [www.jncet.org](http://www.jncet.org), 7(8).

# Cyber Range: Preparing for Crisis or Something Just for Technical People?

Jani Päijänen<sup>1</sup>, Karo Saharinen<sup>1</sup>, Jarno Salonen<sup>2</sup>, Tuomo Sipola<sup>1</sup>, Jan Vykopal<sup>3</sup> and Tero Kokkonen<sup>1</sup>

<sup>1</sup>JAMK University of Applied Sciences, Jyväskylä, Finland

<sup>2</sup>VTT Technical Research Centre of Finland, Tampere, Finland

<sup>3</sup>Masaryk University, Brno, Czech Republic

[jani.paijanen@jamk.fi](mailto:jani.paijanen@jamk.fi)

[karo.saharinen@jamk.fi](mailto:karo.saharinen@jamk.fi)

[jarno.salonen@vtt.fi](mailto:jarno.salonen@vtt.fi)

[tuomo.sipola@jamk.fi](mailto:tuomo.sipola@jamk.fi)

[vykopal@ics.muni.cz](mailto:vykopal@ics.muni.cz)

[tero.kokkonen@jamk.fi](mailto:tero.kokkonen@jamk.fi)

DOI: 10.34190/EWS.21.012

**Abstract:** Digitalization has increased the significance of cybersecurity within the current highly interconnected society. The number and complexity of different cyber-attacks as well as other malicious activities has increased during the last decade and affected the efforts needed to maintain a sufficient level of cyber resilience in organisations. Due to Industry 4.0 and the advanced use of IT and OT technologies and the adaptation of IoT devices, sensors, AI technology, etc., cybersecurity can no longer be considered to be taken lightly when trying to gain a competitive advantage in business. When transferring from traditional reactive cybersecurity measures to proactive cyber resilience, cyber ranges are considered a particularly useful tool for keeping the organisation in the game. With their background in defence research (e.g., DARPA NCP in 2008), cyber ranges are defined as interactive simulated platforms representing networks, systems, tools, and/or applications in a safe, legal environment that can be used for developing cyber skills or testing products and services. Cyber ranges can be considered vital in facilitating and fostering cybersecurity training, certification, and general education. Despite the definition, cyber ranges seem to be only used by military or so-called “technical people” when quite a few more organisations could benefit from them. This article attempts to reveal the secrets behind cyber ranges and their use focusing on suitable target environments, common functions, and use cases. Our main objective is to identify a classification of cyber ranges and skills related to these diverse types of ranges. We emphasise the cyber resilience of any type of organisation that demands the use of cyber range type of training. Different training scenarios improve different sets of organisational skills. The article is based on an extensive survey on cyber ranges, their use, and technical capabilities that was conducted in CyberSec4Europe project.

**Keywords:** cyber range, cyber resilience, cyber training, organisational skills, cybersecurity

---

## 1. Introduction

Given the concept of a cyber crisis (or even cyber war), one has to imagine a cyber weapon being used in a cyber-attack, for example of a malware program or a denial-of-service attack. This attack is usually directed towards a victim (organization or person) that is facing a crisis situation. Different countries have different laws protecting the victim against this kind of aggression. Outside the realm of cyber security, there are usually various kinds of laws prohibiting and restricting the usage of physical weapons, even to the point of having specialized physical shooting ranges abiding the law (Ministry of Interior, Finland, 1998/2003) for the practice of regular weaponry. In the cyber context, these kinds of cyber weapon shooting ranges are being formed as cyber arenas or cyber ranges; however, the development of regulations on how these platforms should be used is currently lacking.

Cyber ranges (or cyber arenas) are technical platforms that facilitate education, training, and exercise of cyber security (Karjalainen and Kokkonen, 2020a). According to Russo, Costa and Armado (2020), these ranges are complex infrastructures that simulate real-world cybersecurity scenarios. These technical platforms have developed in different organizations simultaneously from smaller technical laboratory environments to cloud-based solutions. They might have originally been platforms used to demonstrate products and technology, or even mirroring a technical production network to act as an introductory platform for new employees. Ukwandu et al (2020) have identified current trends, types, target domains and technologies used in cyber ranges and testbeds. On the other hand, the definition of cyber ranges does not limit or restrict use cases, target groups, or participant roles utilizing a cyber range (ECSO, 2020).

## **2. At whom cyber ranges are targeted?**

Cyber ranges can be used for training or educating individuals or groups of people such as employees of companies or organisations. They can be used for cyber security research and development, hosting various kinds of events, certifying products or services, performing competence assessment, or recruiting people (ECSO, 2020; Yamin, Katt, and Gkioulos, 2020). Some cyber ranges can be used to train cyber defence (NATO CCDCOE, 2020; Vykopal et al., 2017). Events in a cyber range can be cyber security exercises or competitions targeted at a company (FINGRID, 2017), an organisation (Valtori, 2020; MITRE, 2014), international (NATO CCDCOE, 2020), or national cyber security exercises (Secretariat of the Security Committee, 2019). An exercise can target a specific audience without any shared training or background (CyberSec4Europe, 2021). Also, various cyber security related competitions such as Capture the Flag (CTF) competitions targeted at individuals or teams can be organised as a cyber range event. Firstly, the following sections introduce target groups benefiting from cyber ranges and secondly, use cases that the cyber ranges have supported.

### **2.1 Target groups**

#### *Individuals, Personal Knowledge, Skills and Abilities (KSA)*

Cyber ranges offer a technical environment where citizens can train their understanding of the cybersecurity phenomena. The European Union has produced the European Qualifications Framework (EQF), which helps to improve transparency, comparability, and portability of people's qualifications between the nations in the EU. These qualifications are listed as learning outcomes Knowledge, Skills and Abilities (KSAs). Cyber ranges could be used in a Cyber Security Massive Open Online Course (MOOC) implementation (Fischer-Hübner et al., 2020), where the MOOCs offer a platform for everyone to improve their KSAs.

#### *Curriculum students*

These KSAs are developed through degree programmes following a curriculum suited for the respected EQF level. Curriculum students of higher education (Karjalainen, Kokkonen and Puuska 2019; Saharinen et al., 2019; Karjalainen and Kokkonen, 2020b) are sometimes required to pass courses that utilize these cyber ranges. Regardless these courses being either a mandatory or elective part of their studies, many education and research organizations are developing the capability (Frank et al., 2017) to host courses through these environments as the demand for capable workforce increases constantly in the field of Cyber Security.

#### *Companies*

Companies invest in protecting their environments, as digitalization is forcing them to be increasingly available online both in the private and public sectors. To uphold these availability requirements, companies need to employ a capable workforce provided by the education sector (Bell and Oudshoorn, 2018). Students with practical knowledge of handling a live cyber crisis are often valued, and the capability of upholding the cyber presence of a company simultaneously with a cyber crisis can be seen as a part of the cyber resilience of a nation.

#### *Law enforcement*

Additionally, individuals face the problem of a cyber crisis when e.g., their digital identity is stolen, or payment frauds are committed in the e-banking realm (Singh and Rastogi, 2018). In both companies and individual cases, these cyber crises end up in police cybercrime statistics. Cybercrimes are investigated by specialized police units that survey and handle cybercrimes for prosecution. Exact methods of cybercrime investigation are still a developing field, which also means the police forces need an educational environment for investigating cybercrimes.

#### *Government*

If the cyber crisis that either faces companies or individuals exceeds a certain threshold, a nation has to implement its laws and regulations to enter a state of war (Sevis and Seker, 2016). This means, depending on the country in question, that the military can start protecting its civilians and assets, be they physical or cyber.

After these laws or regulations are invoked, the protection of assets is commonly left to the nation's military forces.

#### *Military cyber defence capabilities*

The Defence Forces of different countries have been mentioned to use National Cyber Ranges: Norway (NTNU, 2018), Estonia (Republic of Estonia Centre of Defence Investment, 2020) and Finland (JYVSECTEC, 2017; EU2019.fi, 2019) to name a few. Additionally, multinational coalitions have practiced in self-contained cyber ranges brought about for the need, for example, Locked Shields (NATO CCDCOE, 2020). Different military forces have stated that cyber is the fifth domain of warfare after land, sea, air, and space (NATO, 2016).

#### *Researchers*

All the aforementioned entities have Cyber Security researchers (ENISA, 2020a; 2020b; 2020c) working separately and in coalition on different research projects. The development of cyber ranges as such is a less researched area, as the phenomena and results after working in the cyber range are typically more sought after.

## **2.2 Use cases for cyber ranges**

*Security research, testing, development, and certification:* Development testbeds, research environments, and certification tracks have been used in the industry for longer than the term Cyber Range has existed. Development testbeds are usually set up by development teams to see how their updates work in an environment mimicking the production environment. Research environments aim at closeness to the real thing, or a phenomenon is researched by scientists, often relying on ICT environments separated from the Internet. Certification bodies require that the test samples pass through a set of phases on a track in order to gain a label of quality provided by the entity awarding the certificates.

*Security Education through Competence Building and Assessment:* Competence building follows the said certification bodies to offer practicing environments, i.e. cyber ranges, for students trying to reach validation for their skills. This thought has brought up the environment itself to be an active area for student assessment how their competence has developed while working within the environment.

*Development of Cyber Capabilities and Resilience:* The earlier mentioned competence building is a part of an individual's growth as an expert. The development of cyber capabilities and resilience looks at the phenomenon, outcomes using a cyber range, from the organisation's viewpoint, e.g. Fingrid, 2017. One part of it is recruitment, where organizations look for competent workforce, and the interview process might have recruitment sections handled in a technical cyber environment. Additionally, ongoing personnel might be trained using organizational exercises.

*Cross-domain development environment (Digital Dexterity):* The digital dexterity of the whole domain is developed when multiple organisations from multiple industries participate in a cyber range dedicated to the particular industries. These exercises usually show the weak points of processes in multiple organizations, e.g., supply chain processes.

*National and International Cybersecurity Competitions or Exercises:* National or international cybersecurity competitions, in which individuals, organizations or nations compete against one another as well as national and international cyber security exercises, may both advance all the aforementioned use cases.

## **3. Cyber range usage based on a survey**

In this section, we analyse the data from a conducted cyber range survey. The survey was conducted in the CyberSec4Europe project, and it was open from 23 April 2020 to 27 May 2020. A total of 44 responses were received, of which 39 responses were considered valid. The number of survey responses, 39, is considered valid based on the survey authors' experience in the subject. In the survey terms, we decided not to publish any cyber range specific features and capabilities. The survey consisted of single-choice, multiple-choice and open questions, and it did not contain any mandatory fields. (CyberSec4Europe, 2020)

### 3.1 Cyber range target groups

The survey data had a total of seven target groups (TGs) listed, and respondents provided three additional target groups. Hence, the data comprised a total of ten target groups: General public, Secondary level students, Degree program students (Bachelor’s or Master’s degree students), Government organizations, Companies and Enterprises, Non-profit associations or similar, Other, and respondent reported Training Service Providers, Systems Integrators, and Cyber Professionals. The respondents belonged to the following target groups: Training Service Providers, Systems Integrators and Cyber Professionals. They are presented in the columns of Figure 1. The most represented target groups were Companies and Enterprises 77% (30), Degree program students (Bachelor or Master’s degree students) 59% (23), Government organizations 59% (23), Non-profit associations or similar 23% (9), General Public 18% (7) and Secondary level students 18% (7). The following groups were represented in the data by just one respondent: Training Service Providers, Systems Integrators, Cyber Professionals and Other. The top 20% of the cyber ranges supported four or more target groups.

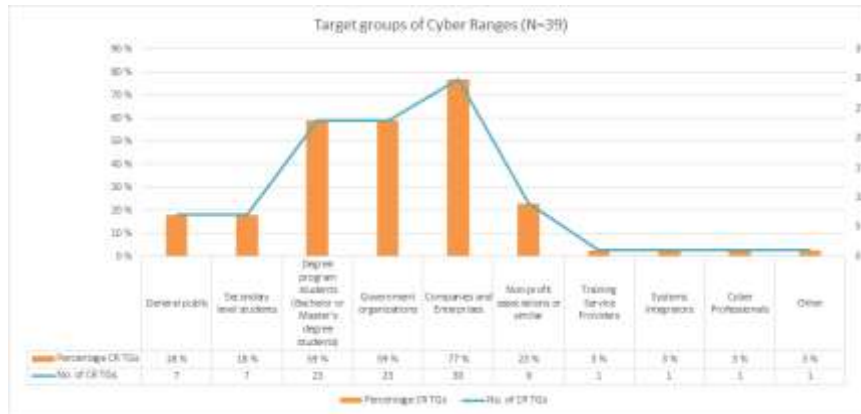


Figure 1: Distribution of target groups (N=39)

The number of target groups supported by cyber ranges is shown in Figure 2. Single Target Group was reported by 23% (9), two target groups by 28% (11), three target groups by 26% (10), four target groups by 13% (5), five target groups by 5% (2), and six target groups by 5% (2). Based on the survey data, a cyber range supports two (2.6) target groups on average.

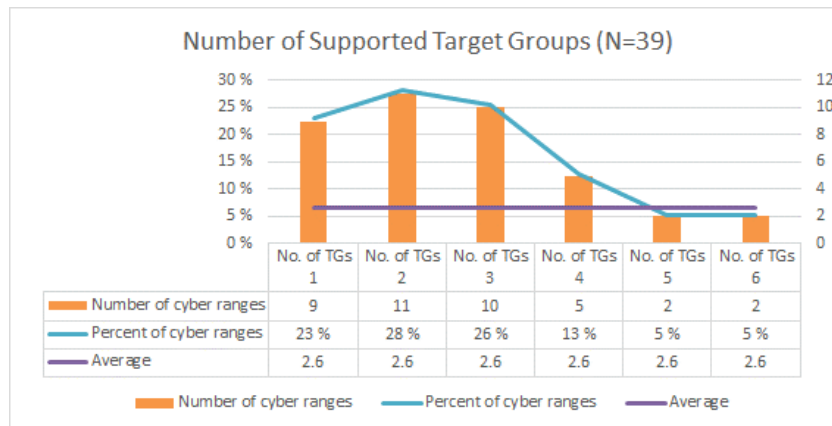


Figure 2: Number of supported target groups (N=39)

### 3.2 Cyber range use cases

A cyber range may be dedicated to a single use case, or it may support multiple use cases. The survey data contained 11 use cases, namely Security testing and certification, Security research & development, Competence Building, Security Education, Development of Cyber Capabilities, Development of Cyber Resilience, Competence Assessment, Recruitment, Cross-domain development environment (Digital dexterity), National and International Cybersecurity Competitions, and National and International Cybersecurity Exercises. The reported use cases were distributed (Figure 3) as Security testing and certification 44% (17), Security research & development 72% (28), Competence Building 62% (24), Security Education 82% (32), Development of Cyber Capabilities 51% (20), Development of Cyber Resilience 38% (15), Competence Assessment 36% (14),

Recruitment 13% (5), Cross-domain development environment (Digital dexterity) 13% (5), National and International Cybersecurity Competitions 26% (10), National and International Cybersecurity Exercises 44% (17).

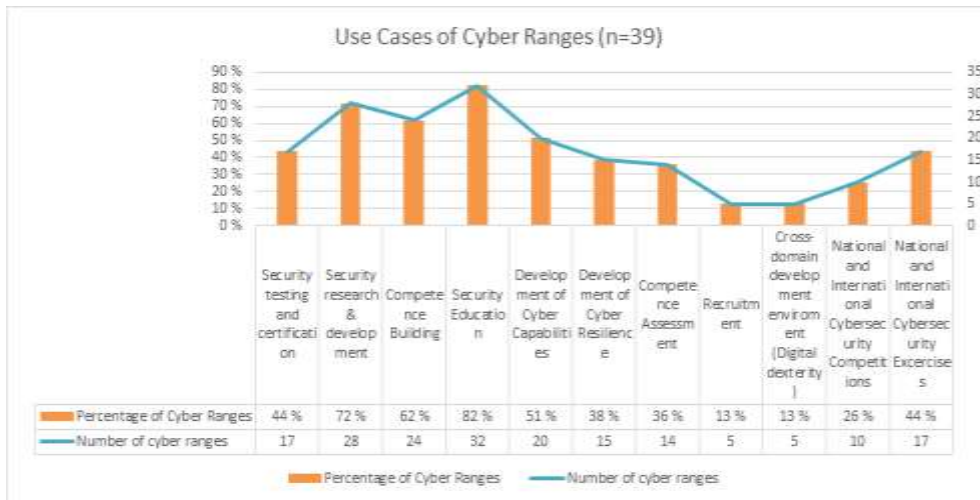


Figure 3: Distribution of use cases (N=39)

Figure 4 displays the number of the use cases (No. of UCs) supported by the cyber ranges. All eleven use cases were supported by 5% (2), ten use cases by 5% (2), nine use cases by 5% (2), eight use cases by 3% (1), seven use cases by 10% (4), six uses cases by 10% (4), five use cases by 10% (4), four use cases by 10% (4), three use cases by 13% (5), two use cases by 13% (5), one use case by 15% (6) cyber ranges as reported by the respondents. On average, a cyber range supports four (4.79) use cases. The top 20% of cyber ranges supported eight or more use cases.

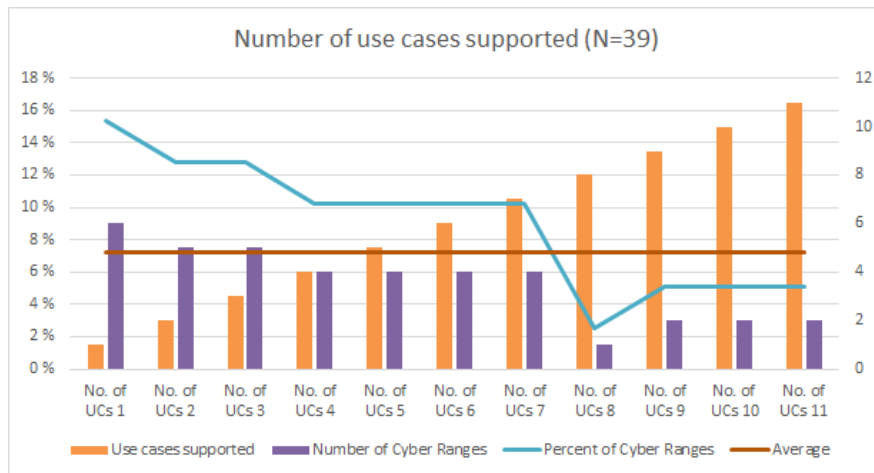


Figure 4: Number of use cases supported by cyber ranges (N=39)

### 3.3 Cyber range participant roles

Six user roles were listed: Director (Business, Director, Communication, etc.), Developer, Researcher, Security professional, Educator, and Other. The survey respondents reported to option “Other” with the following: Sysadmin, Network admin, Student, Job Applicants, Employees, Domain specialist. Two respondents responded “Different roles from organisations which are responsible for some parts of cyber incident response & handling (e.g. Public relations, Process owners, System owners, Technical specialists)” and “CISO, Incident managers, depending on the roles in organisations (e.g. IT admins).”

The number of participant roles is shown in Figure 6: one role 21% (8), two roles 23% (9), three roles 28% (11), four roles 13% (5), and five roles 13% (5). One respondent (3%) did not report the number of participant roles. On average, a cyber range supports two participant roles (2.66%). No cyber ranges were reported to support all roles, including the “Other” role.



Figure 5: Distribution of participant roles

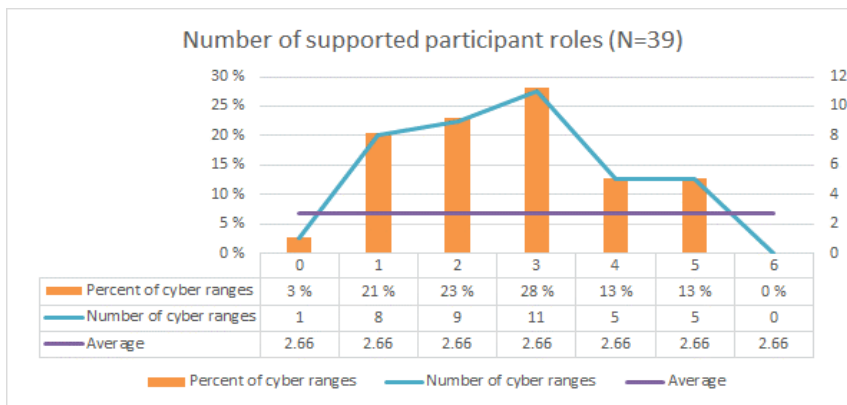


Figure 6: Number of participant roles supported (N=39)

### 3.4 Cross-tabulation of cyber range use cases and participant roles

Table 1 shows the cross-tabulation of filtered data, where target groups were Government organizations, Companies and Enterprises, or Non-profit associations or similar. It shows which use cases a cyber range supports, and the user roles supported. The table rows represent use cases and the columns the user roles. The number following a use case reports the total number of times the use case was reported: Security testing and certification (57), Security research & development (90), Competence Building (75), Security Education (85), Development of Cyber Capabilities (62), Development of Cyber Resilience (50), Competence Assessment (47), Recruitment (20), Cross-domain development environment (Digital dexterity) (21), National and International Cybersecurity Competitions (35), National and International Cybersecurity Exercises (54). In total, the user roles shown in the table were reported as follows: Director (Business, Director, Communication, etc.) 83 times, Developer 94 times, Researcher 145 times, Security professional 161 times, Educator 106 times, and Other User Roles seven times. In each use case the reported cyber ranges supported all the roles, except Other User Roles.

Table 1: Cross-tabulation of use cases with participant roles, filtered.

Use case	Director	Developer	Researcher	Security professional	Educator	Other User Roles	Total
Security testing and certification	8	10	14	14	11	0	57
Security research & development	11	15	22	20	15	7	90
Competence Building	9	12	18	22	14	0	75
Security Education	11	12	21	24	17	0	85
Development of Cyber Capabilities	10	10	15	17	10	0	62
Development of Cyber Resilience	8	8	11	14	9	0	50
Competence Assessment	6	7	11	14	9	0	47
Recruitment	3	3	5	5	4	0	20
Cross-domain development environment (Digital dexterity)	4	4	5	5	3	0	21
National and International Cybersecurity Competitions	4	5	10	10	6	0	35
National and International Cybersecurity Exercises	9	8	13	16	8	0	54
<b>Total</b>	<b>83</b>	<b>94</b>	<b>145</b>	<b>161</b>	<b>106</b>	<b>7</b>	<b>596</b>

## **4. Discussion**

According to the research data, cyber ranges had various target groups (Figure 1), and the supported participant roles of cyber ranges were not limited to technically oriented user roles, but there were roles for e.g., directors (Figure 5). The cyber ranges supporting directors as a potential participant role, support a broader spectrum of use cases (Table 1). The data indicates that cyber ranges were used by both technical and non-technical user roles.

When an entity, e.g., an organisation, a company or an individual faces a cyber incident, it does not require only technical skills to understand, resolve and respond to the incident but also non-technical skills are required (Fingrid, 2017). An organisation may establish a Cyber Security Incident Response Team (CSIRT) that tries to respond to and resolve the attack. According to Onwubiko and Ouazzane (2020), CSIRTs should have the necessary expertise and support from the infrastructure and networking teams, systems administration and management teams, business continuity and disaster recovery teams, communications and press office, and designated senior management teams. In case of severe enough incident, senior management could provide decision-making and funding support; a cyber incident may require a dedicated cost-budget that only the senior management can allocate. The CSIRT example and exercising or training for incidents can be seen as preparing for a local and limited duration crisis. The work to recover from a cyber incident may last long, even several months, depending on the size of the organisation. In larger organisations, the CSIRT team contains these dedicated roles.

In conclusion, the key question of this article “Are cyber ranges just for technical people or do they actually provide vital tools for the organisation to prepare against a crisis?”, we might say that based on our research results, cyber ranges enable the organisations to carry out more than just technical mitigation measures. However, this highly depends on the decisions made by the organisation itself on how well they take the different functionalities into use and make full use of the platform. Simply said, a cyber range acquired only for a specific technical purpose might be somewhat limited in terms of functionality. Since there are quite a few cyber range platforms available on the market with various features ranging from single technical point solutions to comprehensive cyber arenas including realistic simulation of business processes and technical systems, selecting the right tool for a specific organisation might require thorough examination of available options and possibly even external consultation.

The research results show that some cyber ranges support or have participated in national or international cybersecurity exercises. Such exercises, when exercising joint operations of civil government and authorities, or security authorities, require there to be non-technical participants, so that the areas of responsibilities as stated by national or international laws are followed.

Individuals, cyber professionals, government organisations, companies and enterprises, and degree program students use cyber ranges for competence building and development. The business features and domains as well as the technical features and functionalities they provide for users should be researched further. As the original survey was not specifically designed for the purpose of analysing the scope of educational cyber range use, there is a definite need for a new survey. The questions should be adjusted so that their scope focuses more on the previously studied subject and perhaps includes multiple different subjects. Future research might focus on the features, functionalities and properties of cyber ranges which have been reported to support non-technical roles for a better understanding of the potential use cases that they could participate for.

### **Conflict of Interest**

The authors declare no conflict of interest.

### **Acknowledgments**

This research was supported by the Cyber Security Network of Competence Centres for Europe (CyberSec4Europe) project of the Horizon 2020 SU-ICT-03-2018 program, and by the ERDF project “CyberSecurity, CyberCrime and Critical Information Infrastructures Center of Excellence” (No. CZ.02.1.01/0.0/0.0/16\_019/0000822).

The authors would like to thank Ms. Tuula Kotikoski for proofreading the manuscript.



## References

- Bell S. and Oudshoorn M. (2018) "Meeting the Demand: Building a Cybersecurity Degree Program with Limited Resources," 2018 IEEE Frontiers in Education Conference (FIE), San Jose, CA, USA, 2018, pp. 1-7, DOI: 10.1109/FIE.2018.8659341.
- CyberSec4Europe. (2020) "D7.1 Report on existing cyber ranges, requirements", [online], Cyber Security for Europe (CyberSec4Europe), [https://cybersec4europe.eu/wp-content/uploads/2020/09/D7.1-Report-on-existing-cyber-ranges-and-requirement-specification-for-federated-cyber-ranges-v1.0\\_submitted.pdf](https://cybersec4europe.eu/wp-content/uploads/2020/09/D7.1-Report-on-existing-cyber-ranges-and-requirement-specification-for-federated-cyber-ranges-v1.0_submitted.pdf)
- CyberSec4Europe. (2021) "CyberSec4Europe Hosting Flagship 1: An Online Cybersecurity Exercise", [online], Cyber Security for Europe (CyberSec4Europe), <https://cybersec4europe.eu/cybersec4europe-hosting-flagship-1-an-online-cybersecurity-exercise/>
- ECISO. (2020) Understanding Cyber Ranges: From Hype to Reality. [Online]. Available at: <https://ecs-org.eu/documents/publications/5fdb291cdf5e7.pdf>, European Cyber Security Organisation (ECISO), Brussels, Belgium.
- ENISA. (2020a) "ENISA Threat Landscape 2020 - Insider Threat". ISBN:978-92-9204-354-4. DOI:10.2824/552242
- ENISA. (2020b) "ENISA Threat Landscape 2020 - Main incidents", [online], European Union Agency for Network and Information Security (ENISA), Science and Technology Park of Crete (ITE), Heraklion, Greece, Heraklion, Greece, <https://www.enisa.europa.eu/news/enisa-news/enisa-threat-landscape-2020>
- ENISA. (2020c) "ENISA Threat Landscape 2020 - The year in review", [online], European Union Agency for Network and Information Security (ENISA), Science and Technology Park of Crete (ITE), Heraklion, Greece, Heraklion, Greece, <https://www.enisa.europa.eu/news/enisa-news/enisa-threat-landscape-2020>
- FINGRID magazine. (2017) "Cyber security is ensured with genuine exercises, [online], Fingrid Oyj, <https://www.fingridlehti.fi/en/cyber-security-ensured-genuine-exercises/>
- Finnish Ministry of Interior. (2015). "Firearms Act", [Online], <https://www.finlex.fi/fi/laki/kaannokset/1998/en19980001.pdf>.
- EU2019.fi (2019) "Cyber Ranges Federation – Towards Better Cyber Capabilities Through Cooperation" [online], Finnish Presidency of the Council of the European Union (EU2019.fi), <https://eu2019.fi/en/-/cyber-ranges-federation-yhteistyolla-kohti-parempaa-kyberkyvykkytta>
- Fischer-Hübner S. et al. (2020) Quality Criteria for Cyber Security MOOCs. In: Drevin L., Von Solms S., Theocharidou M. (eds) Information Security Education. Information Security in Action. WISE 2020. IFIP Advances in Information and Communication Technology, vol 579. Springer, Cham. [https://doi.org/10.1007/978-3-030-59291-2\\_4](https://doi.org/10.1007/978-3-030-59291-2_4)
- Frank, M., Leitner, M. and Pahi, T. (2017) "Design Considerations for Cyber Security Testbeds: A Case Study on a Cyber Security Testbed for Education," 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Orlando, FL, 2017, pp. 38-46, DOI: 10.1109/DASC-PiCom-DataCom-CyberSciTec.2017.23.
- JYVSECTEC (2017) "JYVSECTEC success story", [online], Jyväskylä Security Technology (JYVSECTEC), <https://jyvsectec.fi/2017/02/jyvsectec-success-story/>
- K. N. Sevis and E. Seker, "Cyber warfare: terms, issues, laws and controversies," 2016 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), London, 2016, pp. 1-9, DOI: 10.1109/CyberSecPODS.2016.7502348.
- Karjalainen, M., Kokkonen T. and Puuska, S. (2019) "Pedagogical Aspects of Cyber Security Exercises", IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Stockholm, Sweden, pp 103–108. DOI: 10.1109/EuroSPW.2019.00018
- Karjalainen, M. and Kokkonen, T. (2020a) "Comprehensive Cyber Arena; The Next Generation Cyber Range", 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Genoa, Italy, pp. 11–16. DOI: 10.1109/EuroSPW51379.2020.00011.
- Karjalainen, M. and Kokkonen, T. (2020b) "Review of Pedagogical Principles of Cyber Security Exercises", Advances in Science, Technology and Engineering Systems Journal, Vol 5, No 5, pp 592–600. DOI: 10.25046/aj050572.
- NATO. (2016) "NATO Cyber Defence" [Online]. Available at: [https://www.nato.int/nato\\_static\\_fl2014/assets/pdf/pdf\\_2016\\_07/20160627\\_1607-factsheet-cyber-defence-eng.pdf](https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2016_07/20160627_1607-factsheet-cyber-defence-eng.pdf)
- NATO CCDCOE. (2020) "Exercises", [Online]. Available at: <https://ccdcoe.org/exercises/>.
- NTNU. 2018. Norwegian University of Science and Technology. "Norwegian Cyber Range". <https://www.ntnu.no/ncr>
- Onwubiko C., Ouazzane, K. (2020) "SOTER: A Playbook for Cybersecurity Incident Management", IEEE Transactions on Engineering Management, DOI: 10.1109/TEM.2020.2979832.
- Republic of Estonia Centre of Defence Investment. 2020. "Estonia Signs Contract to Develop Command Platform for NATO Cyber Range". [Online]. Available at: <https://www.kaitseinvesteeringud.ee/en/estonia-signed-a-contract-for-the-development-of-a-command-platform-for-the-nato-cyber-range/>
- Russo, E., Costa, G, Armado, A. (2020) "Building next generation Cyber Ranges with CRACK", Computers & Security, Vol 95, pp. 101837. DOI: 10.1016/j.cose.2020.101837.
- Singh S. K. and Rastogi N. (2018). "Role of Cyber Cell to Handle Cyber Crime within the Public and Private Sector: An Indian Case Study," 2018 3rd International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Bhimtal, 2018, pp. 1-6, DOI: 10.1109/IoT-SIU.2018.8519884.

- Saharinen K., Karjalainen M., and Kokkonen T. (2019) "A Design Model for a Degree Programme in Cyber Security". In Proceedings of the 2019 11th International Conference on Education Technology and Computers (ICETC 2019). Association for Computing Machinery, New York, NY, USA, 3–7. DOI: 10.1145/3369255.3369266
- Secretariat of the Security Committee. (2019) "Turvallisuusviranomaiset kehittävät osaamistaan kansallisessa kyberturvallisuusharjoituksessa", [Online]. Available at: <https://turvallisuuskomitea.fi/tiedote-turvallisuusviranomaiset-kehittavat-osaamistaan-kansallisessa-kyberturvallisuusharjoituksessa-kyha19-jamkissa-jatkossa-myos-terveydenhuollon-toimijat-mukaan-harjoituksiin/>
- Ukwandu, E., Ben Farah, M.E., Hindy, H., Brosset, D., Kavallieros, D., Atkinson, R., Tachtatzis C., Bures M., Andonovic, I., Bellekens, X. (2020) "A Review of Cyber-Ranges and Test-Beds: Current and Future Trends", arXiv preprint arXiv:2010.06850.
- Valtori. (2020) "Valtori's 2019 financial statements published", [Online]. Available at: [https://valtori.fi/en/-/valtorin-tilinpaaatos-2019-julkaistu?languageId=en\\_US](https://valtori.fi/en/-/valtorin-tilinpaaatos-2019-julkaistu?languageId=en_US)
- Vykopal, J., Ošlejšek, R., Čeleda, P., Vizváry, M., Tovarňák, D. (2017) "KYPO Cyber Range: Design and Use Cases", Proceedings of the 12th International Conference on Software Technologies - Volume 1: ICISOFT, SciTePress, Madrid, Spain, pp. 310-321. DOI: 10.5220/0006428203100321.
- Yamin, M.M., Katt, B. and Gkioulos, V. (2020) "Cyber ranges and security testbeds: Scenarios, functions, tools and architecture", Computers & Security, Vol 88, pp. 101636. DOI: 10.1016/j.cose.2019.101636.

# Multiple-Extortion Ransomware: The Case for Active Cyber Threat Intelligence

Bryson Payne and Edward Mienie

University of North Georgia, Dahlonega, USA

[bryson.payne@ung.edu](mailto:bryson.payne@ung.edu)

[edward.mienie@ung.edu](mailto:edward.mienie@ung.edu)

DOI: 10.34190/EWS.021.075

**Abstract:** In just over three decades since its introduction, ransomware has become a primary security risk to businesses and users, and it is now the fastest-growing category of cybercrime. In addition, ransomware attacks on healthcare, energy and water distribution, and defense contractor organizations have begun to impact both human security and national security. Traditional ransomware encrypts files on an infected computer to block users' access until a sum of money or ransom is paid, often via cryptocurrencies like Bitcoin or Ethereum. Businesses and individuals who fall victim to ransomware are faced with the expense of paying the ransom, restoring files from backup if available, or losing files altogether and starting from scratch. Beginning in late 2019, cybercriminals stepped up their game by deploying new attacks known as "double-extortion" ransomware, wherein files are stolen before being encrypted. Even if an organization might be able to recover its data from backups, by stealing the files first, the attacker can still profit either by selling any confidential data on the dark web or by further extorting the business and threatening to leak sensitive information unless an even larger ransom is paid. As of 2021, double-extortion ransomware is still in its infancy, but the authors anticipate and describe possible long-term trends toward even more persistent multiple-extortion tactics, in which stolen data could continue to be used by cybercriminals, terrorists, and rogue nation-states potentially decades in the future. Traditional, passive measures in cybersecurity and business continuity, like firewalls, antivirus software, and frequent backups, are not sufficient to protect organizations from this new type of data theft and extortion enterprise. Government agencies and private corporations alike are beginning to employ active cyber threat hunters and intelligence analysts to detect and neutralize this newest class of persistent threat. This research examines multiple approaches to more advanced defense against such threats, including the emerging roles of cyber threat hunting and cyber threat intelligence, and the impact of this new type of tradecraft on both current and future multiple-extortion ransomware tactics.

**Keywords:** ransomware, malware, cyberattacks, cybersecurity, cyber threat hunting, cyber threat intelligence

---

## 1. Introduction

For over thirty years, ransomware has locked, hidden, or encrypted files from users in an attempt to extort payment, or ransom, in exchange for a secret key to unlock access to the users' files. Across these past three decades, ransomware has matured into a criminal enterprise costing more than \$20 Billion (USD) per year in damages (Morgan, 2019), affecting governments, businesses, healthcare providers, education, financial institutions, utilities and infrastructure providers, among other organizations (PurpleSec, 2020). Furthermore, ransomware is considered to be the fastest-growing category of cybercrime, increasing by triple-digit percentages annually for the past several years (KnowBe4, 2020).

For much of this time, cybersecurity professionals have advocated for organizations to back up their data regularly, preferably to secure storage located off-site, to minimize the impact of a ransomware attack. With traditional ransomware, an organization might only need to restore their data from the most recent uninfected backup to resume critical operations. However, newer strains of ransomware reported in the past two years have begun to evolve the ability to steal sensitive data from systems before encrypting them, allowing cybercriminals a second opportunity to extort payment from the victim by threatening to release the stolen data to the public or sell it to other criminals on the dark web (O'Donnell, 2020).

This research examines both the emergence and impact of this new type of double-extortion ransomware, as well as methods for defending against and recovering from such aggressive strains of malware. In addition, the authors contribute an analysis and extrapolation of existing and near-term capabilities that portend an even longer tail on such attacks, indicating the evolution of further multiple-extortion ransomware.

## 2. Background

The first ransomware was introduced in 1989—it was distributed by postal mail on a floppy disk disguised as an AIDS research survey, and it encrypted filenames on a user's computer making them essentially unusable after the machine had been rebooted 90 times, demanding a ransom payment of \$189USD (roughly \$400USD

adjusted for inflation to 2021) by cashier's check or international money order to a post office box in Panama (Waddell, 2016). Due to the simple encryption used in the AIDS trojan ransomware, several security researchers were able to reverse engineer the malware and help users recover their files without paying the ransom. Early malware analyst Jim Bates released freely available software within a month of the ransomware's release that decrypted the filesystem and removed the infection from affected systems (Bates, 1990).

While the first reported incidence of ransomware was not regarded as very widespread, advanced, or profitable due to the distribution and payment technology available at the time, it did introduce and popularize the concept of using malware as leverage to extort money from victims (Lessing, 2020). Most prior malware simply destroyed data or inconvenienced the user by consuming all of a computer's resources.

Reports of ransomware generally waned during the late 1990's, and ransomware only began to make a resurgence in 2005 (Liska and Gallo, 2016). But by this time, personal computers were sufficiently powerful that they could perform advanced encryption. And, they were now usually connected to the internet to enable rapid delivery worldwide of ransomware from any location, often by email attachment. With these two advancements, ransomware was able to attack systems more broadly and in a much more sophisticated and devastating manner.

Even with these advancements, most ransomware still extorted relatively small sums per user, partially due to the difficulty in sending payments online anonymously, as credit card companies would reverse fraudulent or criminal charges, and wire transfers could be stopped if reported within 72 hours. But in 2008, the emergence of cryptocurrencies—virtually untraceable, anonymous payment transactions systems like Bitcoin, Ethereum and similar blockchain technologies—breathed new life into ransomware by providing an easy means of collecting ransom from victims over the internet without the normal safeguards of more transparent, legitimate banking systems (Tapsoba, 2018).

### **3. Modern ransomware**

This convergence of high-powered computing devices, universal network access, and anonymous, instant, untraceable payment transfers enabled highly-sophisticated, technologically advanced criminal actors to build well-funded, robust platforms for targeted ransom extortion, often preceded by phishing or other social engineering methods to gain initial access (Boyden, 2020). Today's ransomware has evolved to include advanced encryption, zero-day and one-day attacks, across a wide variety of devices.

Military-grade, asymmetric, RSA encryption is now a staple of most ransomware attacks, securely encrypting a long, unique secret key so that it can only be recovered by paying the ransom and requesting the key from the ransomware distributor. Zero-day attacks are unpatched vulnerabilities usually discovered by security researchers, but these can often be bought over the dark web by well-funded criminal groups and nation-state actors. One-day attacks or n-day attacks are recently patched security vulnerabilities that malware writers seek to take advantage of before users update their systems.

The release of the EternalBlue one-day exploit allegedly developed by the US's National Security Agency (NSA) would contribute to the development of the costliest ransomware attack to date, WannaCry. WannaCry used the leaked EternalBlue exploit to spread rapidly to unpatched Windows computers within a month of Microsoft's release of a security update acknowledging the vulnerability. Cyber risk modelling firm Cyence estimated the total loss caused by WannaCry to be around \$4 billion USD, crippling 200,000 computers in 150 countries including the UK's National Health Service (NHS), who incurred an expense of \$125 million to recover from the attack (Bera, 2021).

In addition, much more than standard desktop and laptop computers can be compromised in a ransomware attack. All computing devices from iPhones to Android devices, tablets to smart TVs can be vulnerable to ransomware techniques (Liska and Gallo, 2016), including smart automobiles (Payne and Abegaz, 2018). In fact, one researcher developed a proof of concept to deploy ransomware to a home's smart thermometer (Franceschi-Bicchierai, 2016), as a demonstration that even so-called Internet-of-Things (IoT) devices could fall prey to ransomware scammers.

In addition to these advanced techniques, the barriers to entry have continued to drop, with the inception of so-called affiliate programs and even Ransomware-as-a-Service (RaaS) operations in which malware developers sell their ransomware to “resellers” who can customize the software easily for a one-time fee or as a subscription service (Auld, 2020). Researchers note that today’s ransomware operations are scalable, enabling even newcomers to build operations generating millions in illicit revenue. The average organization attacked by ransomware today spends more than \$230,000USD to recover, with an average of 19 days’ downtime, enough to threaten or even decimate most small to mid-sized businesses (Bera, 2021).

It should be noted that there are also several risks to organizations who consider paying a ransom hoping to recover their data (DOJ, 2016). First, if the organization does not take sufficient steps to determine how the ransomware infiltrated their systems and remedy those vulnerabilities, the ransomware actors may attack again and again. Second, ransomware attackers may choose to request additional payment after an initial ransom has been paid. Third, as noted by the US Department of Justice’s Cybersecurity Unit, “Paying a ransom does not guarantee an organization will regain access to their data; in fact, some individuals or organizations were never provided with decryption keys after paying a ransom.” (DOJ, 2016, p.5). Or in many cases, a ransomware attacker may not possess sufficient technical skill to successfully retrieve a user’s data even after payment is received due to incompetence or poorly-designed RaaS software. Finally, it is worth mentioning that paying a ransom inadvertently encourages and reinforces this criminal business model.

The most troubling innovation in ransomware in recent history began just about two years ago as a response to users’ refusal to pay to unlock their files because of good backups that could be restored quickly at relatively low cost. Ransomware developers realized that cloud backup services and other low-cost options were enabling more individuals and businesses to recover from even a severe ransomware attack simply by reverting their machine to a clean state and restoring their data from a recent backup. The ransomware writers’ response was both relatively simple and deviously creative: before encrypting a computer, the ransomware would slowly siphon important files and documents from the victim’s filesystem for days or even weeks, sending them in small chunks over the internet to a remote command-and-control server, giving the attacker a copy of some or nearly all of a victim’s data before enacting the ransom. In this way, a victim who refused to pay to unlock their computer could still be pressured into paying to keep their stolen data from being released.

#### **4. Double-Extortion ransomware**

It is the data stealing capability of newer ransomware reported since 2019, most notably Maze, Ryuk, and DoppelPaymer, that has given rise to the new double-extortion tactic as part of a multi-phased attack (Trend Micro Research, 2021). In double-extortion ransomware attacks, the threat actor first infiltrates a high-value network to steal data, quietly exfiltrating data from victim systems for a period of time before engaging the encryption phase of the attack to lock the victim’s files. If the victim does not pay the ransom required to unlock their files after phase two, the threat actor then threatens to release copies of sensitive data stolen from the victim through an aptly-named data leak site.

The first publicly disclosed double-ransomware attack by the Maze criminal group demanded roughly \$2.3 million (USD) in bitcoin as ransom from a security and facilities services company to decrypt their network. When the company refused to pay, the ransom was increased by 50% to almost \$3.8 million and a 700-megabyte file containing sensitive company data was leaked publicly to pressure the company into paying the higher ransom to avoid damage from the release of further sensitive information. Within two years of Maze’s first publicized double-extortion ransomware attack, researchers discovered a dozen and a half similar data leak sites from peer and competitor ransomware actors (Auld, 2020).

If the victim chooses not to pay to have their files decrypted, the cybercriminals running the ransomware ring may first choose to threaten to leak sensitive data by linking to a small sample of the victim’s stolen documents, such as a client list, transaction records, an employee database, or patients’ medical records. This makes it much more likely that a victim will pay, as a sensitive data breach is usually much more damaging than simply losing access to files for a limited time (Hero, 2020). Most data breaches will require notification of individuals whose information may have been compromised, as well as notification of law enforcement relevant government agencies as required by law, and some types of data breaches may cause legal action and civil or criminal penalties to be levied against the ransomware victim, as in the case of personal medical information protected by HIPAA (the Health Insurance Portability and Accountability Act), personal data of EU citizens and residents

covered by GDPR (the European Union's General Data Protection Regulation), among other protected types of data.

In addition to economic security, national security is equally imperilled by these new threats. In one case, a US defense contractor was held for double-ransom early in 2020 after having data stolen by the Ryuk Stealer module of the Ryuk ransomware (Cimpanu, 2020). The increased effectiveness and higher ransom amounts made possible by double-extortion ransomware has fuelled speculation that these new tactics will expand even further in the next few years.

The level of opportunism demonstrated by some ransomware actors has also become a threat to human security. Unfortunately, even in the midst of a global pandemic, ransomware attackers took advantage of the rush to find a vaccine and the crush on local hospitals to steal files and extract double-ransoms from hospitals (Winder, 2020), research centers and healthcare providers (Gallagher, 2020).

## **5. The future of ransomware**

Double-extortion ransomware is still in its relative infancy—as of this writing, barely two years have passed since the first such attacks. While the one-two punch of newer double-extortion ransomware is disconcerting, it seems to presage an even longer tail on such data-theft and ransomware combination attacks. At present, many cybercrime rings are using data leak sites to post sensitive files from victims unwilling or unable to pay ransom amounts ranging from several thousand to several million dollars. But in addition to the first and second ransom demand, cybercriminals may be opening the door to further extortion attempts by leaking or selling stolen information.

The authors propose a new term, multiple-extortion ransomware, to describe ransomware attacks that are designed give a threat actor multiple opportunities for extracting payment from their victims. Current strains of double-extortion ransomware that steal data have already opened the door for further extortion down the road, but they have not yet shown the intent to attempt later blackmail beyond the second ransom attempt. Multiple-extortion ransomware refers to data-theft-capable ransomware that extracts sensitive information, encrypts a user's or organization's files, and intentionally applies multiple, staged extortion techniques, either against the same user or organization or against customers, employees, or other third parties whose information may have been contained in the original data theft and ransomware attack.

There are many ways in which these kinds of multiple-extortion attacks may already be developing organically, with or without forethought. First, third parties who purchase or download a victim's sensitive files may pursue additional attempts to blackmail or otherwise extort money from the victim directly using the same information. Second, criminals may leverage customer, client, patient, or employee data from leaked files to harass, intimidate, or extort money from those individuals. Third, and possibly most concerning of all, while an individual or organization might not be a valuable or high-profile target at the present time, the negligible cost of storing stolen data could mean that many years in the future, the sensitive information of a potential world leader or prominent businessperson could be used to blackmail or extort years or possibly decades after an initial attack.

This type of long-game strategy may be beyond the interest or grasp of many smaller cybercriminals, but nation-state actors who already engage in decades-long (or centuries-long) rivalries are well-accustomed to such long-game tactics, and indeed may already be storing or utilizing stolen, ransomed data in such a manner. Even many larger cybercrime rings are showing evidence of longer-term thinking, with many ransomware variants waiting 30 days, 90 days, or even longer to perform the encryption phase of the attack, thus ensuring that multiple backup copies across several weeks or months are infected and therefore likely vulnerable to re-attack even after an organization restores systems from a first attack. In fact, the authors have first-hand knowledge of a ransomware attack against a local business in which the company was forced to restore from backup every day for almost a month until a threat-hunting team was finally able to detect and remove all copies of the ransomware from across hundreds of system that had remained hidden for more than a month through dozens of backup cycles.

However, many indicators point to more intentional types of combination, multiple-extortion attacks. There have been numerous documented incidents involving industrial cyber-physical attacks, in which cyberattacks leverage computer systems to shut down electric power or damage mechanical or manufacturing equipment

(Mienie and Payne, 2019). Consider the possibility of an advanced criminal organization or nation-state actor infiltrating a critical infrastructure network and first stealing data, then implanting destructive software capable of cyber-physical damage, waiting a predetermined period (say, several months) to ensure that even the organization's backup files contain the malware, then launching a three-phase ransomware attack in which 1) the organization's files are encrypted, risking downtime or loss of business, 2) sensitive files are leaked over the internet, demanding a further ransom to recover or destroy the stolen data, and 3) a final ransom attempt threatens to shut down critical systems or cause physical damage. Or, perhaps a smart automobile is first disabled requiring ransom payment or a hardware reset of the main control unit to re-enable the automobile, then camera footage is released from the car showing that the attacker has personal information on the driver, then, as a last resort, the attacker attempts to crash the vehicle while in motion. All of these techniques exist individually, but taken together, a multiple-extortion ransomware attack would be difficult to fend off and devastating to recover from for individuals and organizations alike.

## **6. Active cyber threat intelligence**

Given the complexity of the threat, the cybersecurity industry is adopting and adapting tactics, techniques, and procedures (TTPs) from the intelligence community in fighting these new advances in cybercrime, cyberwarfare, and multiple-extortion ransomware. In addition to the multiple layers of traditional controls like antivirus software on devices, intrusion detection and prevention systems (IDS/IPS), whitelisting or regulating the software that can run on a system, and a regular regimen of updating software on at most a weekly cycle, organizations are employing cyber threat intelligence in new ways.

In traditional government intelligence collection and analysis, and in particular counterintelligence efforts, information on nefarious actors is gathered and a concerted effort is made through focused activities to first identify the target, then deceive them, followed by their exploitation and disruption with the ultimate aim of protecting the nation's national security interests. Counterintelligence is in part a collection issue and it is more than a security or law enforcement issue as highlighted by the three types of counterintelligence efforts conducted by the government. First, collection with the aim of obtaining information about the opponent's intelligence gathering capabilities. Second, by assuming a defensive posture, efforts are made to thwart the hostile actor's efforts to penetrate the government's agencies. Third, once the opponent's efforts to penetrate the nation's own systems are identified, an effort is made to manipulate these nefarious attacks by turning the hostile actors into double agents or by supplying them with false information (Lowenthal, 2017).

And in cyber teams throughout organizations, threat intelligence is being used to anticipate, detect, prevent, and remedy cyberattacks like ransomware. First, organizations can use industry, government, and third-party threat intelligence from common vulnerabilities and exposures (CVE) reports, information sharing and analysis centers (ISACs), open-source, and subscription-based threat reports to inform their preparedness efforts to anticipate near-term attacks based on reports from peers and the industry as a whole. Second, organizations are employing more active cyber threat hunters (an actual job title of its own these days), including malware analysts and reverse engineers to take apart malware samples and figure out how they work and how to stop them, threat intelligence analysts to comb through threat reports gleaning best practices to share with their cyber teams, and other innovative roles combined with artificial intelligence-based tools to detect patterns and anomalies. These human-plus-machine teams are key elements in detecting and preventing especially novel threats like zero-day and one-day attacks that may not yet appear in antivirus signature databases.

## **7. Conclusions and recommendations**

In addition to backing up files regularly and using both host-based antivirus/anti-malware and network-based intrusion detection systems, organizations must defend more actively against evolving threats by keeping systems updated and patched to ensure known vulnerabilities are remedied, as well as by considering data loss prevention tools to detect and prevent the exfiltration of data that occurs early in a double-extortion or multiple-extortion ransomware attack (Fearn, 2020).

It should be noted that threat hunting and active cyber threat intelligence are not the only tools in an organization's or in a nation's arsenal for defeating ransomware. In addition to advanced AI tools and well-trained human cyber operators, organizations may engage in civil and criminal proceedings against threat groups, and nation-states can make use of diplomatic and economic sanctions in retaliation for cyberattacks (Tapsoba, 2018). In addition, many organizations are acquiring cyber insurance to cover losses in the event of a

ransomware attack, with 41% of all cyber insurance claims in the first half of 2020 being attributed to ransomware (Raghunarayan, 2020).

A multi-layered, holistic cyber strategy employing both active and passive controls, technology, training, and human-machine teaming is becoming the standard for mid-sized and larger organizations. The need for specialized cyber threat hunters skilled in threat intelligence analysis, malware analysts trained in reverse engineering, and security teams embedded across multiple business units and throughout IT who understand the importance of mission-critical operations and staying ahead of threat actors may seem like a tall order. But the future of our national security, economic security, and even human security demand that we adapt to the constantly evolving nature of cyber threats and train up a new generation of cyber heroes to battle multiple-extortion ransomware and other near-term cyberattacks.

## References

- Auld, A. (2020). What's behind the increase in ransomware attacks this year? PwC Cyber Threat Operations. Retrieved from <https://www.pwc.co.uk/issues/cyber-security-services/insights/what-is-behind-ransomware-attacks-increase.html>
- Bates, J. (1990). Trojan horse: AIDS information introductory diskette version 2.0. *Virus Bulletin*, 3-6.
- Bera, A. (2021). 22 Shocking Ransomware Statistics for Cybersecurity in 2021. SafeAtLast.co. Retrieved from <https://safeatlast.co/blog/ransomware-statistics/>
- Boyden, P. (2020, November 30). Double extortion – Ransomware at another level. FraudWatch International. Retrieved from <https://fraudwatchinternational.com/all/double-extortion-ransomware-at-another-level/>
- Cimpanu, C. (2020, January 29). DOD contractor suffers ransomware infection. ZDNet. Retrieved from <https://www.zdnet.com/article/dod-contractor-suffers-ransomware-infection/>
- Department of Justice Cybersecurity Unit. (2016). How to Protect Your Networks from Ransomware. June 2016. Retrieved from <https://www.justice.gov/criminal-ccips/file/872771/download>
- Fearn, N. (2020, August 27). Double extortion ransomware attacks and how to stop them. Computer Weekly.
- Franceschi-Bicchierai, L. (2016). Hackers makes the first-ever ransomware for smart thermostats. Motherboard.com. Retrieved from <https://www.vice.com/en/article/aekj9j/internet-of-things-ransomware-smart-thermostat>
- Gallagher R. (2020, April 1). Hackers 'without conscience' demand ransom from dozens of hospitals and labs working on coronavirus. Bloomberg. Retrieved from <https://fortune.com/2020/04/01/hackers-ransomware-hospitals-labs-coronavirus>
- Hero. (2020). What you need to know about double extortion ransomware attacks. Hero Managed Services. Retrieved from <https://www.heromanaged.com/what-you-need-to-know-about-double-extortion-ransomware-attacks/>
- KnowBe4. (2020). Ransomware Timeline. KnowBe4, Inc. Retrieved from <https://www.knowbe4.com/ransomware#ransomwaretimeline>
- Lessing, M. (2020, June 3). Case Study: AIDS Trojan Ransomware. SDX Central. Retrieved from <https://www.sdxcentral.com/security/definitions/case-study-aids-trojan-ransomware/>
- Liska, A., & Gallo, T. (2016). *Ransomware: Defending against digital extortion*. O'Reilly Media, Inc.
- Lowenthal, Mark M. (2017). *Intelligence, from Secrets to Policy (7th edition)*. Los Angeles, U.S.: CQ Press.
- Mienie, E.L., Payne, B.R. (2019). The Impact of Cyber-Physical Warfare on Global Human Security. *International Journal of Cyber Warfare and Terrorism (IJCWT)* 9 (3). pp. 36-50.
- Morgan, S. (2019, October 21). Global Ransomware Damage Costs Predicted To Reach \$20 Billion (USD) By 2021. Cybercrime Magazine. Retrieved from <https://cybersecurityventures.com/global-ransomware-damage-costs-predicted-to-reach-20-billion-usd-by-2021/>
- O'Donnell, L. (2020, April 16). 'Double Extortion' Ransomware Attacks Spike. ThreatPost. Retrieved from <https://threatpost.com/double-extortion-ransomware-attacks-spike/154818/>
- Payne, B. R., & Abegaz, T. T. (2018). Securing the Internet of Things: best practices for deploying IoT devices. In *Computer and Network Security Essentials* (pp. 493-506). Springer, Cham.
- PurpleSec. (2020). 2020 Ransomware Statistics, Data, & Trends. Retrieved from <https://purplesec.us/resources/cyber-security-statistics/ransomware/>
- Raghunarayan, R. (2020, December 1). Are you prepared for double extortion attacks? SANS. Retrieved from <https://www.sans.org/blog/are-you-prepared-for-double-extortion-attacks-/>
- Tapsoba, K. (2018). *Ransomware: Offensive Warfare Using Cryptography as a Weapon* (Doctoral dissertation, Utica College).
- Trend Micro Research. (2021, January 5). An Overview of the DoppelPaymer Ransomware. Trend Micro. Retrieved from [https://www.trendmicro.com/en\\_us/research/21/a/an-overview-of-the-doppelpaymer-ransomware.html](https://www.trendmicro.com/en_us/research/21/a/an-overview-of-the-doppelpaymer-ransomware.html)
- Waddell, K. (2016, May 10). The computer virus that haunted early aids researchers. *The Atlantic*, 1-2. Retrieved from <https://www.theatlantic.com/technology/archive/2016/05/the-computer-virus-that-haunted-early-aids-researchers/481965/>
- Winder, D. (2020, April 16). *Hospitals on COVID-19 frontline facing 'double extortion' cyber threat*. Forbes. Retrieved from <https://www.forbes.com/sites/daveywinder/2020/04/16/hospitals-on-covid-19-frontline-face-double-extortion-threat-security-experts-caution>



# Resilience Management Concept for Railways and Metro Cyber-Physical Systems

Jyri Rajamäki

Laurea University of Applied Sciences, Espoo, Finland

[jyri.rajamaki@laurea.fi](mailto:jyri.rajamaki@laurea.fi)

DOI: 10.34190/EWS.21.074

**Abstract:** Railways and metros are good examples of cyber-physical systems (CPS). They are safe, efficient, reliable and environmentally friendly. However, being such critical infrastructures turns metro, railway and related intermodal transport operators into attractive targets for cyber and/or physical attacks. SAFETY4RAILS H2020 project of the European Commission delivers methods and systems to increase the safety and resilience of track-based inter-city railway and intra-city metro transportation. Safety engineers have established strategies over decades to remove risks and increase safety that become manifest in railway systems. On the other hand, resilience is a multi-faced and not yet standardized concept so that a number of definitions and assessment methods exist, and until now, resilience management has largely focused on descriptive or diagnostic analytics following an expert judgment-based approach. This paper aims at introducing a conceptualization for resilience management of CPS and to bring the lessons to be learned from earlier projects for SAFETY4RAILS. The approach, earlier studied in the healthcare sector, is based on an integration of the concept of cyber-trust with cybersecurity science and resilience science. The paper proposes five principles that arise from the theory for resilience management processes of CPS: (1) design and implement a security management plan, (2) employ all appropriate security technologies, (3) ensure the adequacy and quality of security information, (4) make sure that situational awareness is always up to date, and (5) design and implement a resilience management plan that covers all four event management cycles (plan/prepare, absorb, recovery, adapt) and interdependencies with other systems. In addition, the paper discusses the meaning of these principles in the rail transportation sector. The paper represents the author's views having taken part in SAFETY4RAILS stakeholder workshops as part of the stakeholder needs and requirements analysis in the early stages of the project.

**Keywords:** cybersecurity, resilience management, cyber-physical systems, SAFETY4RAILS project, rail transportation systems

---

## 1. Introduction

Railway systems and metros are relatively safe, efficient, reliable, comfortable and environmentally friendly mass carriers. Nowadays, they have become even more important from a sustainable point of view since the consequences of the climate change have become more evident. However, being part of the critical infrastructure (CI) makes railway and metro operators and transportation operations that depend on safe railways and tracks attractive targets for cyber and/or physical attacks, which poses a security threat. SAFETY4RAILS H2020 project idea concentrates on a more reliable safety and security of inter-city transport (railways) and intracity transport (metros) that depend on railways and tracks. Until now, past incidents in Europe refer only to cyber-attacks (e.g. WannaCry) or physical attacks (e.g. the bombing attacks to the commuter trains in Madrid in 2004) and not to combined cyber-physical attacks. In the event of a combined cyber-physical attack, the reaction (response and mitigation) will need to take into account several aspects and to be structured to limit the casualties, serious transport disruptions and significant financial and economic losses. In the European Railways Infrastructures, cyber-attacks on industrial control systems increased by more than 600% between 2012 and 2014. Railway specifics, such as electronic components scattered along tracks or trains, a very long life cycle (in excess of 25 years), diversity both of supply chain and technology and other characteristics make this a complex and challenging domain from both cyber and physical security perspective (European Commission, 2020).

Safety engineers have established strategies over decades to remove risks and increase safety that become manifest in railway systems. On the other hand, resilience is a multi-faced and not yet standardized concept so that a number of definitions and assessment methods exist, and until now, resilience management has largely focused on descriptive (i.e., what happened) or diagnostic analytics (i.e., why it did happen) following an expert judgment-based approach (Bellini, et al., 2021). This paper aims at introducing a conceptualization for resilience management of cyber-physical systems (CPS), and discusses the meaning of that in the rail transportation sector. SAFETY4RAILS is well-connected to already existing and funded projects, and the target of this paper is to bring the knowledge and lessons to be learned for SAFETY4RAILS from earlier projects, such as SHAPES (c.f. Rajamäki, 2020) and ECHO (c.f. Rajamäki & Katos, 2020). According to SAFETY4RAILS stakeholder workshops as part of the stakeholder needs and requirements analysis in the early stages of the project, resilience challenges of railway

and metro systems are quite similar than the ones of, for example, in the healthcare sector, both being critical cyber-physical systems having wide variety of Internet of things (IoT) sensors.

The knowledge base of this study consists of (1) concepts of trust-building in the digital world, (2) resilience science, and (3) cybersecurity science. The paper deduces five principles for cybersecurity and resilience management of cyber-physical systems from the theory. The rest of the paper is structured as follows: Section 2 presents the themes of cyber-trust and cyber resilience. Section 3 deduces a cybersecurity model for a cyber-physical system combining the concept of cyber-trust with cybersecurity science. Section 4 develops the model and presents the resilience management framework and the five principles. Section 5 discusses the results in the rail transportation sector, and Section 6 concludes the paper.

## 2. Cyber-trust

Investing in systems that improve confidence and trust can significantly reduce costs and improve the speed of interaction. From this perspective, cybersecurity is a key enabler for the development and maintenance of trust in the digital world, and the overall goal of cybersecurity is that all operational systems and infrastructures are resilient. According to DIMECC (2017), cybersecurity has the following four themes: (1) security technologies, (2) cognitive situational awareness, (3) security management, and (4) cyber resilience (resilience of operational systems), as shown in Figure 1.

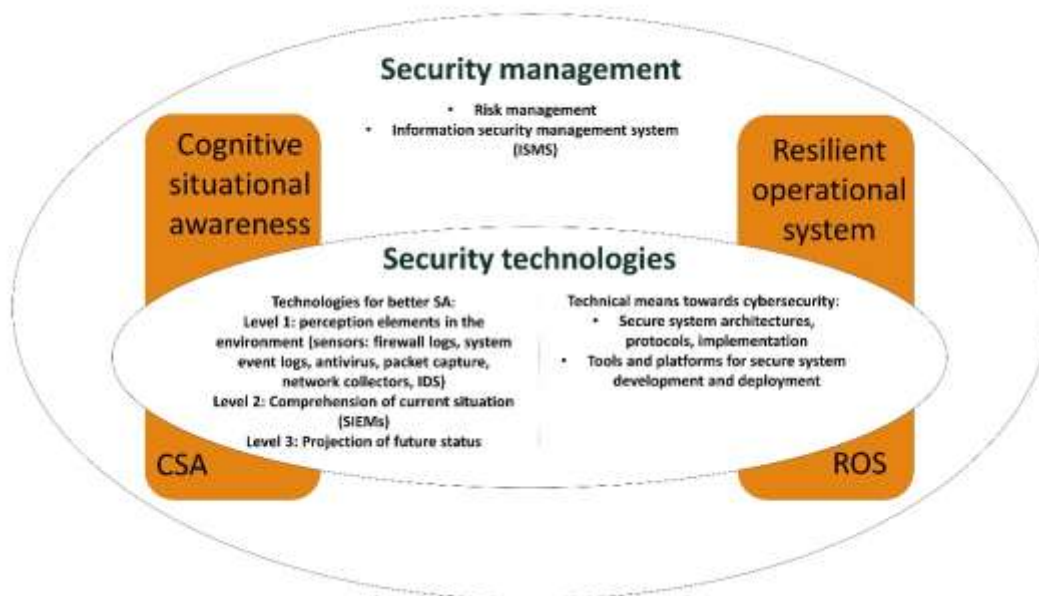


Figure 1: Themes of trust-building in the digital world (modified from DIMECC, 2017)

### 2.1 Security management

Security management focuses on the continuous management and operation of a system by the documented and systematic establishment of the procedures and processes to achieve confidentiality, integrity, and availability of the organization's information assets that do the preservation. Its focus areas include security policy development and implementation, risk management and information security investment, incentives, and trade-offs. From an organization's point of view, cybersecurity management starts by a risk management procedure. If cybersecurity risks are not prepared for, organizations will face severe disasters over time. Risk management research focuses on how to measure and quantify a state of cybersecurity, including quantifying the value of cybersecurity to an operation, how much of a threat is the operation exposed to, and scoring how mitigations and security controls affect the overall operational risk (Edgar & Manz, 2017). All organizations are becoming more and more dependent on unpredictable cybersecurity risks. Everywhere present computing means that organizations do not know when they are using dependable devices or services and there are chain reactions of unpredictable risks. An information security management system (ISMS) provides controls to protect organizations' most fundamental asset, information. Many organizations apply audits and certification for their ISMS to convince their stakeholders that the security of the organization is properly managed and meets

regulatory security requirements. An information security audit is an audit on the level of information security in an organization. Security aware customers may require ISMS certification before a business relationship is established. Now, the most common information technology (IT) security standards are ISO/IEC 27001:2013 Information security management systems requirements, ISO/IEC 27002:2013 Code of practice for information security controls, and ISO/IEC 27005:2018 Information security risk management. Unfortunately, IT security standards are not perfect and they possess potential problems. Usually, guidelines are developed using generic or universal models that may not apply to all organizations. Guidelines based on common, traditional practices take into consideration differences of the organizations and organization-specific security requirements.

## **2.2 Security technologies**

Security technologies include all technical means towards cybersecurity, such as secure system architectures, protocols, and implementation, as well as tools and platforms for secure system development and deployment. Security technologies enable the technical protection of infrastructures, platforms, devices, services, and data. Technical protection starts with secure user identification and authorization that are necessary features in most secure infrastructures, platforms, devices, and services. Fortunately, well-known technologies exist for their implementation. Typically, processes and data objects are associated with an owner, represented in the computer system by a user account, who sets the access rights for others. Well-known security technology standards are ISO/IEC 27033:2015 Network security, ISO/IEC 27034:2015 Application security, and ISO/IEC 27036-1:2014 Information security for supplier relationships.

Technologies that create or transfer security information from the resilient operational system (ROS) to the cognitive situational awareness (CSA) system include sensors that collect the first level of data. Commonly, host- and network-based tools generate logs that are used for CSA. Firewalls, system event logs, antivirus software, packet captures, net flow collectors, and intrusion detection systems are examples of common cyberspace sensors (Edgar & Manz, 2017). Level-two technologies generate information from the data to determine a current situation. Generally, level-two technologies require the bringing together of data and performing some level of analytics. The simplest form is signature-based tools such as antivirus and intrusion detection systems. These systems have encapsulated previous knowledge of detected attacks into signatures that detect and alert when attacks are detected in operational systems. More advanced systems such as security information and event managers (SIEMs) provide infrastructure to bring together datasets from multiple sensors for performing correlations. Vulnerability analysis to determine how many unpatched vulnerabilities exist in a system is also a form of level-two technology (Edgar & Manz, 2017). The third and final level is hard to achieve and only a few examples of effective tools exist. Cyber-threat intelligence provides information on active threat actor methods, techniques, and targets providing some level of predictive information to enable taking pre-emptive security measures (Edgar & Manz, 2017).

Security technologies are needed also when something has happened. For example, forensics can lead to the sources of the attack/mistake and provide information for legal and other ramifications of the issue. Forensics also facilitates the analysis of the causes of the incident, which in turn, makes it possible to learn and avoid similar attacks in the future.

## **2.3 Cognitive situational awareness**

Cognitive situational awareness is the main prerequisite of cybersecurity and resilience. Without CSA, it is impossible to systematically prevent, identify, and protect the system from cyber incidents and if a cyber-attack happens, to recover from the attack. CSA involves being aware of what is happening around your system to understand how information, events, and how your actions affect the goals and objectives, both now and soon. It also enables the selection of effective and efficient countermeasures, and thus, protects the system from varying threats and attacks (DIMECC, 2017). Eckhart, Ekelhart & Weippl (2019) present a CSA framework for CPS based on digital twins that provides a profound, holistic, and current view on the cyber situation. CSA is needed for creating a sound basis for the development and utilization of countermeasures (controls), where resilience focuses. The most important enablers of CSA are observations, analysis, visualization, governmental cyber-policy, and national and international cooperation. For the related decision-making, relevant information collected from different sources of the cyber environment or cyberspace, e.g., networks, risk trends, and operational parameters, are needed.

## 2.4 Cyber resilience

The concept of cyber resilience is similar than the general definition of resilience in other disciplines (Ligo, Kott & Linkov, 2021). Its one general definition includes four abilities: plan or prepare for, absorb, recover from, and adapt to known threats (National Academy of Sciences, 2012). Linkov et al. (2013) combine those abilities with cyber system metrics on four domains (Alberts, 2002), as shown in Figure 2: (1) physical resources and the capabilities and the design of those resources, (2) information and information development about the physical domain, (3) cognitive use of the information and physical domains to make decisions, and (4) the social structure and communication for making cognitive decisions. Cyber resilience of a system can be appreciated only when adequate resilience measures are defined and implemented (Ligo et al., 2021). The process of building resilience is a collective action of public and private stakeholders responding to infrastructure disruptions (Heinimann & Hatsector, 2017).

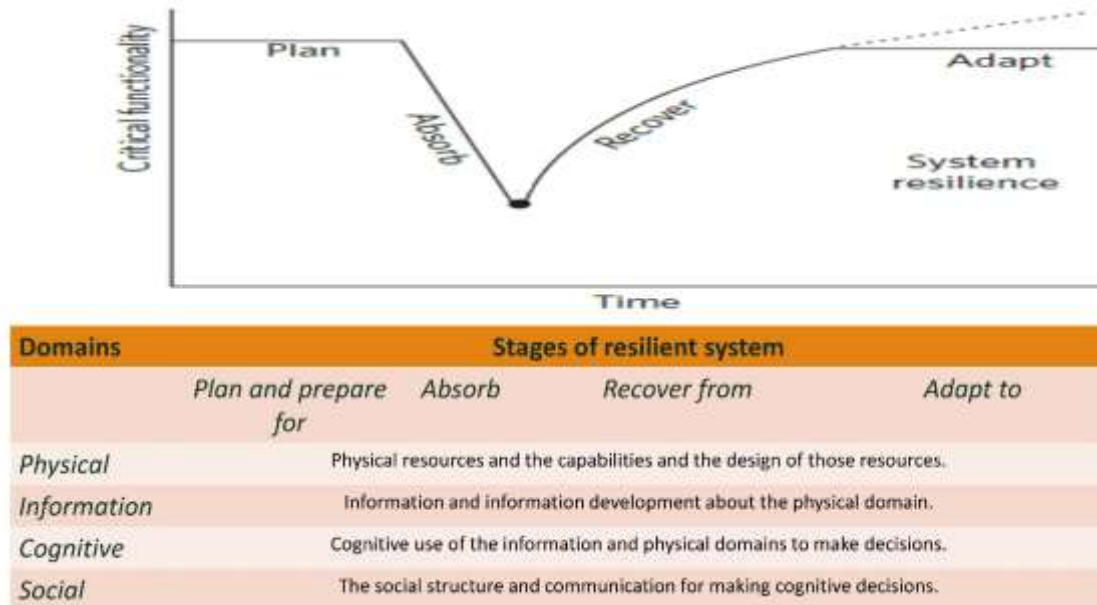


Figure 2: The cyber resilience matrix (modified from Linkov et al., 2013)

## 3. Cybersecurity model for a cyber-physical system

### 3.1 Cybersecurity science

Figure 3 shows three perspectives (domains) of cyberspace: (1) a data or information perspective that comes from the information theory space; (2) a technology perspective that includes the hardware, silicon, and wires, as well as software, operating systems, and network protocols, and (3) a human perspective that acknowledges that the human is as responsible for the dynamics of the system as the data and the technology are (Edgar & Manz, 2017).



Figure 3: Cyberspace at the overlap of data, technology, and human (Edgar & Manz, 2017)

### 3.2 Integration of cybersecurity science with the concept of cyber-trust

All cyber-physical systems have human, technology, and data domains. Previous Figure 1 shows the themes of a resilient CPS consisting of two sub-systems: CSA and ROS. Both of these sub-systems have human (social), technological (physical), and data (information) domains as illustrated in Figure 4. Security management, security technologies, and security information connect these sub-systems together. Security management covers the human and organizational aspects of cybersecurity. Security management also integrates the social layer's operational and cognitive aspects; all organizational and social components should learn from prior events and incidents. Security information is mostly created and/or transferred from ROS to CSA via security technologies. However, the CSA system requires information outside ROS, as shown in the lower part of Figure 4.

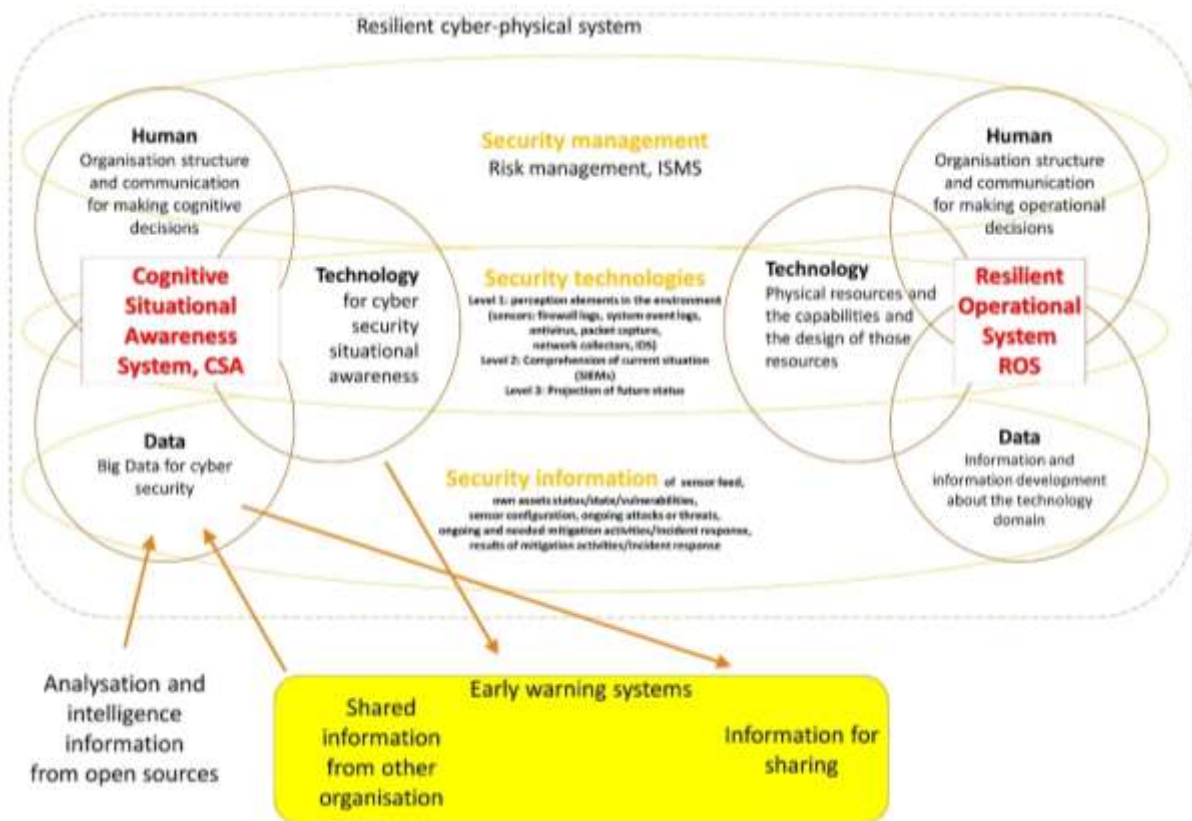


Figure 4: Cybersecurity conceptual model for a cyber-physical system

### 3.3 Cognitive decision-making process

The technology domain of the CSA system includes the data fusion engine, information interfaces and the human-machine interface (HMI) providing an effective visualization layer (Kokkonen, 2016). Cognitive decision-making functionalities should utilize artificial intelligence and be as automatic as possible without human interaction. However, there should be an operator for controlling the sensors and data fusion algorithms and inputting additional information into the system. The system implements HMI for effective visualization of the status of the cyber domain under control and for the input of information that cannot be entered automatically. Humans are needed for controlling the data fusion process and making decisions. HMI should implement different visualizations for different levels of users: e.g. a technical user who requires detailed technical information, whereas a decision-maker needs different visualization. HMI also implements filters for data allowed for different users.

The cognitive situational awareness system in Figure 4 utilizes the information from the operational system to make decisions that aim towards better resilience. The cognitive decision-making process utilizes cybersecurity sensor information as well as the status information of all the known cyber entities. Information on systems, devices, and sensors with their status and configuration information, but also data from used spare parts of physical devices are relevant information for CSA system (Kokkonen, 2016). In addition, information about the

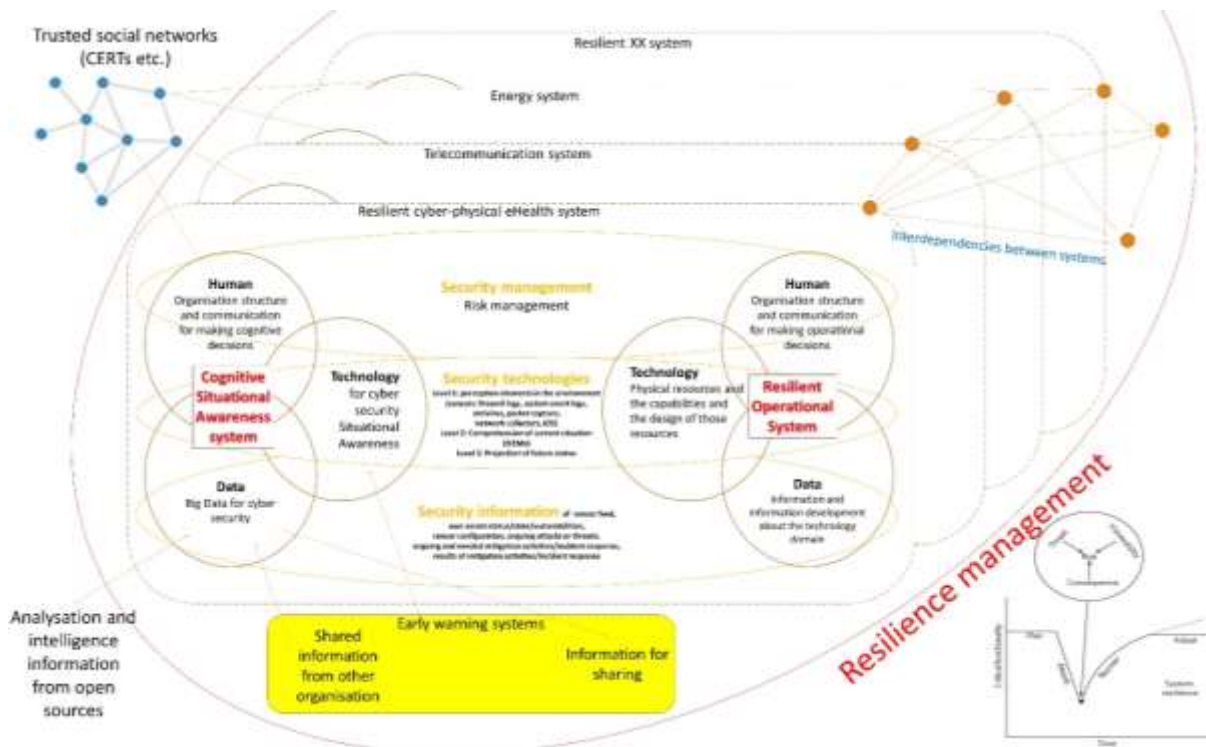


status of saved data and the status of information flows should be reported. Some of that information can be automatically generated using data interfaces and some should be user-generated by using a HMI.

The cognitive decision-making process requires also information exchange between different stakeholders as well as data from open sources, as shown in the lower part of Figure 4. An early warning systems implement interfaces for cybersecurity information exchange with trusted companions. In addition, CSA needs analysis and intelligence information from open sources. That kind of information includes analysed impact assessment information, Indicator of Compromise (IOC) information, and early-warning information from open-source intelligence using, e.g., social media or Computer Emergency Response Team (CERT) bulletins. Further, required policies and objectives should be input into the system.

#### 4. Resilience management framework

Figure 5 presents the conceptual resilience management framework for a resilient rail transportation cyber-system. It is based on the cybersecurity model of a CPS (Figure 4) and interconnections with other CPSs such as telecommunications and energy. Figure 5 presents six important cybersecurity and resilience themes: resilient operational system, security management, security technology, security information, cognitive situational awareness and resilience management. The goal of the management framework is the resilience of the operational system, and that will be achieved via the five other themes. Next, we will give principles how to deal with these themes.



**Figure 5:** Conceptual resilience management framework for a resilient rail transportation system

*Principle 1: Design and implement a security management plan.* This will include the following sub-tasks: carry out cyber risk management; identify and coordinate with external entities that may influence or be influenced by internal cyber-attacks (establish a point of contact); educate and train employees about cybersecurity and the organization’s security management plan; delegate all assets and services to specific employees; prepare security communications; and establish a cyber-aware culture.

*Principle 2: Employ all appropriate security technologies.* This will include the following sub-tasks: implement controls/sensors for critical assets; implement controls/sensors for critical services; assess network structure and interconnection to system components and the environment; implement redundancy of critical physical infrastructure; and assess the redundancy of data physically or logically separated from the network.

*Principle 3: Ensure the adequacy and quality of security information.* This information should be suitability for artificial intelligence and machine learning technologies. This guideline will include the following sub-tasks: categorize assets and services based on the sensitivity; document certifications, qualifications, and pedigree of critical hardware and/or software providers; prepare plans for storage and containment of classified or sensitive information; and identify internal system dependencies.

*Principle 4: Make sure that situational awareness is always up to date.* This cognitive domain will include the following sub-tasks: anticipate and plan for system states and events; understand the performance trade-offs of organizational goals; set up scenario-based cyber war gaming; utilize applicable plans for system state when available; and utilize artificial intelligence or prepare to utilize it for responding to threats with greater confidence and speed.

*Principle 5: Design and implement a resilience management plan that covers all four event management cycles (plan/prepare, absorb, recovery, adapt) and interdependencies with other systems.* This will include the following sub-tasks: consider how all previous requirements can be utilized throughout the four event-management cycles; identify external system dependencies (i.e., telecommunication, energy, built environment), and plan the coordination framework with these systems (you have no control for these systems); and educate and train employees about resilience and the organization's resilience plan.

## **5. Resilience management of railway and metro systems**

The complexity of railway and metro systems combined with existing and emerging cyber-physical threats constrains administrations to consider smart technologies and related huge amounts of data generated as a means to take timely and informed decisions. Therefore, the measures to be put in place for the protection, safeguard and resilience of their infrastructure need to be continuously tested and improved, without forgetting the security of railway workers and passengers (European Commission, 2020). Transportation systems need to be prepared for both expected and unexpected situations and the possibility to mitigate the effect of the uncertainty behind the causes of disruptions through the analysis of all the possible data generated by smart cities open new possibility for resilience operationalization (Bellini, et al., 2021).

### **5.1 Security management plan**

In Europe, the legal background for security management of railway systems comes from the NIS Directive 2016/1148. Its main topics include (1) national capabilities; Member States must have a national Computer Security Incident Response Team (CSIRT), perform cyber exercises, etc., (2) cross border collaboration; cross border collaboration between EU countries, (3) supervision of critical sectors (energy, transport, water, health, digital infrastructure and finance sector); Operators of Essential Services (OES), and critical digital service providers. According to NIS, railway undertakings and infrastructure managers are OES. On the other hand, metros are not included, but it would be convenient to apply NIS to them as well. NIS does not enforce detailed requirements for OES, but the NIS Coordination Group defines the security measures for OES. ENISA maps out the security measures for the railway sector from the most spread standards (ISO 27001 and IEC 62443 Industrial Network and System Security). The proposal of NIS2 Directive has been issued; it has a more detailed framework with respect to NIS, but still it gives no detailed requirements nor mandatory standards.

The security management plan will enrich past natural, cyber and physical events and serve as a basis for identifying the challenges and the respective requirements. The plan will analyse threats, attack vectors and vulnerabilities of the rail-dependant infrastructures in terms of business criticality and support the definition of risk mitigation strategies at planning, protection and prevention level. The mitigation strategies may be either proactive, e.g. resulting in building more robust railway infrastructures and forecasting of events; or reactive, e.g. informing simultaneously the agencies responsible to tackle the threats and its consequences (e.g. railway and metro operators, transportation stakeholders, law enforcement authorities, medical assistance, fire alarms, etc.). In addition, crisis mitigation strategies will be provided, e.g. arranging alternative transportation, rerouting and micro-response activities on different levels, e.g. single stations or at individual passenger level.

### **5.2 Security technologies of railway and metro systems**

Development and design of a resilient cyber-physical rail transportation system starts with the overall system architecture, which describes all systems and environments and then get the systems to work together in a

controlled way with cybersecurity and information security in mind. In addition, all relevant cybersecurity technologies discussed in the section 2.2 should be applied. Common standardised physical security technologies relevant for railway and metro systems include, e.g.:

- CENELEC EN 50132 7:2012 CCTV surveillance systems for use in security applications
- DD CLC/TS 50131 7:2010 Alarm systems Intrusion and hold up systems. Application guidelines
- OASIS Common Alerting Protocol (CAP)
- Open Network Video Interface Forum (ONVIF) Core.

### **5.3 Security information related to railway and metro systems**

The data applied in the CSA system includes all the data generated within the transportation system; operational status information, sensor data from physical railway network (e.g. CCTV, gas sensors in metro systems, metal detectors) and sensor data from railway IT infrastructure. Bellini, et al. (2021) present a concept how to utilize urban big multimedia data (U-BMD) for operationalising the resilience of transport systems. U-BMD is very diverse in terms of volume, velocity, and variety, as well as in terms of accessibility and license for reuse. Useful U-BMD for CSA includes geographic information system maps (seismic risk maps, hydrological risk maps, services, descriptors of structures such as schools, hospitals, infrastructures, etc.), social media streams, IoT-data streams, CCTV streams, etc.

### **5.4 Cognitive situational awareness of railway and metro systems**

The focus of SAFETY4RAILS is to develop a flexible multi-lingual SAFETY4RAILS Information System (S4RIS) that can be combined with already existing safety and security control systems of the railway operators. This combination will be an AI-oriented detection, mitigation, prevention, forecasting and fast response CSA system. S4RIS will analyse the impact of proposed strategies in both the prevention and response phases. S4RIS will combine simulation and monitoring capabilities as well as visualisation means to prevent, forecast, detect, defuse, respond and mitigate the impact of cyber and physical threats in a holistic methodological and operational approach resulting in a collaboration between cyber-physical security technologies and actors. The simulation capabilities will simulate the current practices of the considered railway infrastructure, identify potential vulnerabilities and cascading effects due to various incidents, test/stress the results of the designed security measures and propose alternatives for handling the critical points of the infrastructure.

### **5.5 Resilience management plan**

Trains and metros play a crucial role for both inter-city and intra-city transportation as both may have cascading effects influencing the effectiveness of mobility of people not only between cities but also within cities. Resilience is not about the performance of individual railway or metro system elements but rather the emerging behaviour associated to intra and inter system interactions. Nowadays our societies provide a wide range of transportation and other services available to citizens in real-time, regardless of location or time. This also means that almost all systems are interconnected through different integration platforms. There are also many federations between information systems. Before we can design resilient cyber-physical systems, we need to look at different scenarios, use cases, and requirements. In addition, large cross-system integrations and federations in ICT systems mean that there is a lot of interdependence between different ICT systems and between organizations and stakeholders. We need to identify dependencies on all internal and external systems and data flows, and only then, we can design and implement resilience in cyber-physical systems.

SAFETY4RAILS aims at addressing cyber-physical railway threats resulting in business disruptions, causing time-consuming and even fatal consequences through innovative solutions consisting not only of rerouting approaches but also of mobile infrastructure components like mobile stations, mobile bridges or mobile signal systems in order to react more efficiently and to save time and recovery costs. On the other hand, Bellini, et al., 2021 propose the following multi-steps approach for the resilience management of urban transport systems:

- 1. Understanding the transportation system and utilizing the Functional Resonance Accident Model (FRAM) in managing critical events
- 2. Understanding what information is needed to take decisions
- 3. Selecting/producing U-BDM: methodologies to be adopted to select and collect the data needed



- 4. U-BDM collection and integration: data collection
- 5. U-BDM sense making, how the data is transformed into information
- 6. Knowledge driven decision: how the information is transformed into knowledge.

## 6. Conclusions

This paper offers a conceptual resilience management framework for resilient cyber-physical systems and presents five principles for resilience management. The first principle ‘design and implement a security management plan’ is based on the long security management tradition considering that we are safe and we must plan how to protect ourselves from the outside coming threats by risk management. The second principle ‘employ all appropriate security technologies’ continues this tradition giving tools for protection. The third principle ‘ensure the adequacy and quality of security information’ means that you need data for understanding your system, understanding informational needs for decision making and selecting/producing data needed to support such decisions. The fourth principle ‘make sure that situational awareness is always up to date’ means that you should transform above-mentioned data into knowledge that supports decision-making, and the future evolution of the CSA system may be represented by the digital twin (c.f. Bellini, et al., 2021). The last principle ‘design and implement a resilience management plan that covers all four event management cycles (plan/prepare, absorb, recovery, adapt) and interdependencies with other systems’ goes beyond risk-based security management recognizing that no one can control and protect the whole system of infrastructures when grave incidents, such as the COVID-19 pandemic, happens in any case. Then, you should know your most critical assets and do everything to keep them in the life. These five principles above are deduced from the theory, and so, they are intended for permanent use. On the other hand, the descriptions of the principles and the discussion about them in the case of rail transport systems are based on the state of the art, and these descriptions and discussion will become out-of-date with the progress of technology.

## Acknowledgements

Acknowledgement is paid to SAFETY4RAILS Project. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 883532. The sole responsibility for the content of this paper lies with the author. It does not necessarily reflect the opinion of the European Commission or of the full project. The European Commission is not responsible for any use that may be made of the information contained therein.

## References

- Alberts, D. (2002). Information age transformation, getting to a 21st century military. DOD Command and Control Research Program.
- Bellini, E. et al., 2021. An IoE and Big Multimedia Data Approach for Urban Transport System Resilience Management in Smart Cities. *Sensors*, 21(435), pp. 1-34.
- DIMECC. (2017). The Finnish Cyber-trust Program 2015–2017. Helsinki: DIMECC.
- M. Eckhart, A. Ekelhart and E. Weippl, "Enhancing Cyber Situational Awareness for Cyber-Physical Systems through Digital Twins," 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Zaragoza, Spain, 2019, pp. 1222-1225, doi: 10.1109/ETFA.2019.8869197.
- Edgar, T., & Manz, D. (2017). *Research methods for cybersecurity*. Cambridge: Syngress.
- European Commission. (2020). Grant agreement number 883532 – SAFETY4RAILS.
- Heinimann, H., & Hatsector, K. (2017). Infrastructure Resilience Assessment, Management and Governance – State and Perspectives. In I. Linkov, J.M. Palma-Oliveira (eds.), *Resilience and Risk*, NATO Science for Peace and Security Series C: Environmental Security (pp. 147-187). Cham: Springer.
- Kokkonen, T. (2016). Anomaly-Based Online Intrusion Detection System as a Sensor for Cybersecurity Situational Awareness System. Jyväskylä studies in computing 251. University of Jyväskylä.
- Ligo, A., Kott, A. & Linkov, I. (2021) How to Measure Cyber Resilience of an Autonomous Agent: Approaches and Challenges, In AICA 2021, 1st International Conference on Autonomous Intelligent Cyber-defence Agents. Paris, France.
- Linkov, I., Eisenberg, D., Plourde, K., Seager, T., Allen, J., & Kott, J. (2013). Resilience metrics for cyber systems. *Environ Syst Decis*.
- National Academy of Sciences. (2012). Disaster resilience: a national imperative.
- Rajamäki, J., 2020. Resilience Management Framework for Critical Information Infrastructure: Designing the Level of Trust that Encourages the Exchange of Health Data. *Information & Security*, 47(1), pp. 91-108.
- Rajamäki, J. & Katos, V., 2020. Information Sharing Models for Early Warning Systems of Cybersecurity Intelligence. *Information & Security*, 46(2), pp. 198-214.

# Digital Evidence in Disciplinary Hearings: Perspectives From South Africa

Trishana Ramluckan<sup>1,2</sup>, Brett van Niekerk<sup>1</sup> and Harold Patrick<sup>1</sup>

<sup>1</sup>University of KwaZulu-Natal, South Africa

<sup>2</sup>Educor Holdings, South Africa

[ramluckant@ukzn.ac.za](mailto:ramluckant@ukzn.ac.za)

[vanniekerkb@ukzn.ac.za](mailto:vanniekerkb@ukzn.ac.za)

[patrick@ukzn.ac.za](mailto:patrick@ukzn.ac.za)

DOI: 10.34190/EWS.21.024

**Abstract:** The prevalence of digital communications results in the need for presenting digital evidence in legal and disciplinary hearings. Whilst South African legislation does provide some guidelines for the use of digital evidence, labour laws do not explicitly consider digital evidence and its role in disciplinary hearings is not explicitly guided. This may result in improper use of digital evidence during disciplinary hearings, which may result in unfair labour practices. This paper considers South African legislation and best practice documents to identify challenges of digital evidence and provide recommendations for improving organisational use of digital evidence in internal proceedings.

**Keywords:** digital evidence, forensic investigation, labour law, digital media preservation, standards

---

## 1. Introduction

In South Africa the Electronic Communications and Transmissions Act (2002) defines the use of digital evidence; however, the labour laws do not explicitly consider the use of digital evidence. Whilst cases that appear before a court will have more formal procedures for the submission of evidence, often disciplinary hearings are less formal and the chairperson may not have a legal background nor experience with digital evidence. Due to the ease with which digital evidence can be tampered with, or incomplete evidence can be presented, there remains a high probability of the occurrence of unfair hearings in cases where corrupted evidence is allowed. Therefore, there needs to be a minimum requirement for the collection and presentation of digital evidence in disciplinary hearings in order to minimise the chance of unfair hearings.

The paper will consider the prevailing legislation in South Africa and relevant best practice documents to propose a set of guidelines for the minimum rigour needed to collect, analyse, store, and present digital evidence in disciplinary hearings. In addition, this paper would also provide insights of acquiring the media (evidence) from the internal target mailbox server for use in the forensic investigation. Timeframes, privacy and integrity of the evidence is of a concern. The paper will highlight the volatility and preservation of the digital evidence. If the process is not forensically sound this can also impact the outcome of the disciplinary hearing. These guidelines are intended to provide organisations with a benchmark for the required capability to internally deal with digital evidence, and for chairpersons of disciplinary hearings to assess if presented digital evidence should be allowed and the strength to afford it during the hearing. Jordaan and Bradshaw (2015) indicate that incorrect convictions in courts can often be attributed to inadequate skills of the forensic practitioners, and that the qualifications of the digital forensic practitioners is below that of global norms. Digital forensics is an evolving field (Mushtaque, Ahsan & Umer, 2015); therefore, there is a need for continuous training of practitioners.

Section 2 presents the prevailing South African legislation related to the law of evidence with a focus on digital evidence, and labour law. Section 3 discusses best practices for digital evidence based on international standards. The challenges of digital evidence are discussed in Section 4, which is followed by a proposed framework for digital evidence in disciplinary hearings in Section 5 and the conclusion in Section 6.

## 2. South African law

### 2.1 Labour law

While South African Labour law does not directly provision for the admissibility of digital evidence, in a recent case which was heard at the South African Labour court of South Africa, WhatsApp messages and videos, were permitted. These videos and messages were used earlier in a disciplinary hearing in which the accused was acquitted after the Passenger Rail Agency of South Africa found that the messages were sent in error.

Nevertheless, of the circumstances, the digital evidence was deemed admissible, firstly at the disciplinary hearing and thereafter at the Labour court (The Labour Court of South Africa, Port Elizabeth, 2018). Arguments arise here as to the admissibility of digital evidence and whether original digital evidence (e.g. original emails or social media) should be explicitly considered in the Labour Regulations.

## **2.2 Technology and privacy law**

### *2.2.1 The Electronic Communications and Transactions Act (ECTA)*

The ECT Act (or Electronic Communications and Transactions Act 25 of 2002) became law in South Africa in 2002. With reference to Michalsons (2008), the ECT Act remains as just one of sources of law which impacts on electronic communications and transactions in South Africa and must be read in conjunction with the applicable statutory and common law. The ECTA is applicable to all types of communication by e-mail, the Internet, SMS etc. There is an exception for the admissibility of voice communication between two parties depending on the circumstances.

Furthermore, the ECT Act is regarded as a broad piece of legislation that provisions for issues which are not entirely related to electronic communications and transactions- examples include the use of cyber inspectors, as well as the liability of service providers and domain names. The ECTA also attempted to provision for legal certainty in vague areas in law e.g. 'click wrap' agreements (Michalsons, 2008). The ECTA is considered as an extremely important piece of legislation and is discussed further in Section 2.3 concerning the admissibility of digital evidence.

### *2.2.2 The Protection of Personal Information Act (POPIA)*

The Protection of Personal Information Act or POPIA was introduced in 2013 as South Africa's data protection law. POPIA aimed at protecting personal information processed by public and private bodies but only came into effect on 1 July 2020, including a 12-month compliance period. POPIA is an important privacy legislation as it includes both the technical and legal processes (Michalsons, 2021). With reference to the admissibility of digital evidence at disciplinary hearings, POPIA accommodates for anonymity of the parties involved. Therefore, the legal procedures initiated by organisations must comply to the privacy regulations i.e. POPIA to maintain the privacy of parties involved.

### *2.2.3 The Regulation of Interception of Communications and Provision of Communication Related Information Act (RICA)*

The Regulation of Interception of Communications and Provision of Communication Related Information Act (RICA) 70 of 2002 took effect on 30 September 2005 in South Africa. RICA's purpose is to ensure the proper governance of the interception or monitoring of paper-based and electronic communications. However, concerning digital evidence and its admissibility in disciplinary hearings and courts of law, organisations would experience difficulties in archiving and using information due to the RICA, as the Act states that all forms of monitoring and interception of communications are unlawful. There are some exceptions to this clause which are:

- Section 4 of the RICA allows a party to a communication to monitor and intercept the communication if he/she is a party to the communication (for example, where the participants in a meeting consent to the meeting being recorded). This exception also applies where the interceptor is acting with the consent of one of the parties to the communication.
- Section 5 permits for the interception of any communication under any circumstances – i.e. no special motivation or reason is required for it provided the person whose communication is being intercepted has consented to it in writing prior to such interception.
- Section 6 applies to businesses. It involves the interception of "indirect communications in connection with the carrying on of business". Section 6 authorises any person to intercept indirect communications in the course of carrying out their business by means of which a transaction is concluded in the course of that business, which "otherwise relates to that business" or which "otherwise takes place in the course of the carrying on of that business, in the course of its transmission over a telecommunication system".

Sections of RICA have since been deemed unconstitutional (Sonjica, 2021), but still remain enforceable.

### **2.3 Law of evidence and digital evidence**

In South Africa, as in the United States, digital data is often referred to as hearsay evidence (Krotoski, 2011; Swales, 2018). While it may be argued that if the data provided is proved to be authentic, then it should not be considered as hearsay or circumstantial evidence, however, by its nature of being malleable and/or ephemeral makes it hearsay by legal definition. While the South African law of evidence, does not fully address the admissibility of hearsay evidence, with digital evidence falling under this ambit, S3(4) of the Law of Evidence Amendment Act (1988), does provision for the admissibility of hearsay evidence and states,

‘(4) For the purposes of this section—

- (a) “hearsay evidence” means evidence, whether oral or in writing, the probative value of which depends upon the credibility of any person other than the person giving such evidence.
- (b) “party” means the accused or party against whom hearsay evidence is to be adduced, including the prosecution.

In cases of the admissibility of digital evidence in South African legislation, whether it pertains to documents from an electronic source, emails, fax, SMS and/or social media are currently being regulated by the Electronic Communications and Transactions Act 25 of 2002 (the ECTA) (Nortje, 2016). The ECTA provides the comprehensive requirements that need to be achieved/reached prior to data messages / electronic documents being admissible as evidence before the court. The communications medium, the originator, addressor, recipient/addressee, require verification under oath and the compliance of such requirements have been provisioned for in the ECTA to be admissible in court.

With reference to S (14) of the ECTA, the data message must meet the legal requirements of being presented or retained in its original form if the following have been achieved -

- a) the integrity of the information from the time it was first generated in its final form as a data message has passed an assessment, on whether the information has remained complete and unaltered in the purpose for which information was generated, with having regard for all circumstances”; and
- b) “that the information is capable of being displayed or produced to the person to whom it is to be presented.”

Sections 15(1) and (2) refer to the admissibility of digital evidence – usually addressed by the Chair, while Section 15(3) takes into consideration the evidence and the strength thereof. This may include the following considerations:

- the reliability of the way the digital evidence was generated, stored, or communicated.
- the reliability of the way the integrity of the digital evidence was maintained.
- the way the originator of the digital evidence was established.
- any other relevant factors.

What must be noted is that documents can be altered, either intentionally or unintentionally, such as by viruses on the computer, or machine, or either intentionally by gaining external access to a computer and changing the material contents of a data message. As in the case of *Ndlovu v Minister of Correctional Services* and another 2006 All SA 165 (W) page 172, it was held that the ECTA does not render data messages admissible without further ado. The main reason being that ECTA prohibits the exclusion from evidence of a data message on the grounds that it was generated by a computer and not a natural person. It is important to note that computer generated messages and content may be vulnerable to alterations and in order for it to be admissible evidence, it must be proven that it is accurate and has not been tampered with.

### **3. Digital evidence best practice**

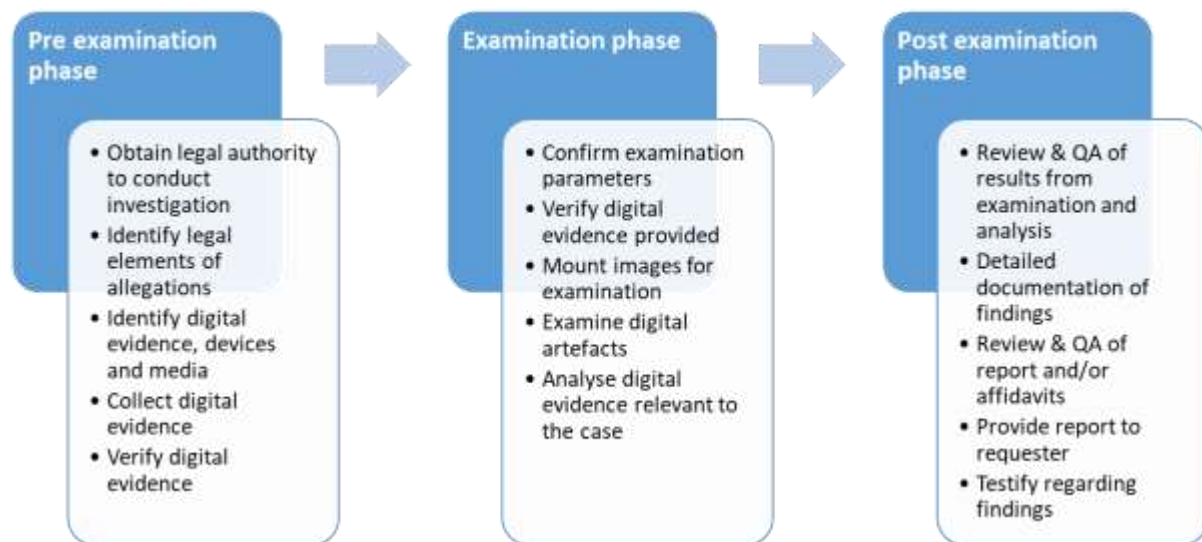
In this section, we will consider some best practice documents to illustrate important aspects of the identification, collection, analysis, storage, communication, and documentation of digital evidence to be used in disciplinary hearings and legal proceedings. Of particular importance will be two South African National Standards (SANS) documents: SANS 27037:2014, based on the ISO/IEC 27037:2012 Information technology — Security techniques — Guidelines for identification, collection, acquisition, and preservation of digital evidence (South African Bureau of Standards (SABS), 2014); and SANS 27035:2013 based on the ISO/IEC 27035:2011

Information technology — Security techniques — Information security incident management document (SABS, 2013). In addition, the ISACA (2015) *Overview of Digital Forensics* and Jordaan's (2016) presentation *Understanding Digital Forensics - Standards, Processes, and the Law* provide a useful overview of the necessary processes.

Digital evidence is usually associated with legal proceedings such as court cases related to cyber-crime (Ali, 2012; Sabillon, Serra-Ruiz, Cavaller, & Cano, 2017). However, the use of digital evidence can also arise for internal disciplinary hearings. Often, these events can be considered similar to a security incident. In the SANS 27035:2013 document, it clearly includes the use of electronic evidence for internal disciplinary proceedings (SABS, 2013: 4; 9; 27). A high-level process for digital evidence is as follows (ISACA, 2015: 9):

- 1. Obtain the necessary authority for searches.
- 2. Document the chain of custody.
- 3. Duplicate digital evidence and validate the copies using hash functions.
- 4. Validate forensic tools to ensure their correct functioning.
- 5. Analyse the collected evidence using appropriate investigative and analytical techniques.
- 6. Repeat and reproduce forensic analysis procedures and conclusions.
- 7. Report the analytical procedures and conclusions.
- 8. Present expert testimony about the findings and conclusions.

Jordaan (2016) divides the process into three phases, as is shown in Figure 1.



**Figure 1:** Digital evidence lifecycle, adapted from Jordaan (2016)

### 3.1 Digital forensic analysts

ISACA (2015: 10) indicates the role of the forensic analyst is to “provide facts and impart knowledge to give expert opinion only when they are required to do so in court. They never seek to aid or blame. Instead, analysts provide a scientific basis so that the court, company or other requesting party may use the unbiased evidence and gain a better understanding of events.”

Section 6.4 of the SANS ISO/IEC 27037 (SABS, 2014) considers the competency of the digital forensic analysts. In particular, the analysts need to “be properly and adequately trained to handle digital devices” (p. 13) for investigation purposes, these skills need to be maintained and demonstrated, and that “it is the responsibility of individual(s) and the employer to ensure that they are adequately trained and the skills and competency maintained” (p. 13). Within the context of disciplinary hearings, the onus of the employer to ensure the adequate skills of those involved in collecting digital evidence is interesting. Disciplinary hearings are not necessarily or rigorous or formal as court proceedings, therefore there is a possibility that evidence could be

collected by an unskilled person (such as email evidence from a colleague or line manager). This may create some liability for the employer should this evidence be found to have been modified or misrepresented.

### **3.2 Documentation and the chain of custody**

Strong documentation is essential as this provides a clear description of all processes and events with respect to the evidence, which provides an indication of the strength and integrity of the evidence. Both the SANS ISO/IEC 27037 and SANS ISO/IEC 27035 documents provide a description of what should be documented (SABS, 2013; 2014):

- Every action or activity related to the evidence;
- Everything that was observed related to the evidence;
- *The date and time settings of any powered-on device, and this should be compared with a reliable date and time source;*
- *Anything visible on the screen of any device (active software, documents and processes);*
- *Any unique identifiers related to the digital devices (or parts thereof) that are of relevance or interest;*
- The location of the digital evidence;
- Any movement or transfer of the evidence;
- Details about the archiving of the evidence;
- How evidence verification was performed;
- The dates, time, and location for all of the above; and,
- The chain of custody.

It should be noted that where any device needs to be accessed, that this should only be done by qualified people who can limit the modification of the system (SABS, 2014).

The chain of custody is the documentation specifically for evidence handling, and its purpose is to mitigate concerns over mishandling, tampering, or misconduct related to evidence (ISACA, 2015; SABS, 2014). Section 6.1 of the SANS ISO/IEC 27037 indicates that a minimum the chain of custody should include (SABS, 2014):

- A unique evidence identifier;
- Logging of access to the evidence (person, time, and location);
- Logging the checking in and checking out of evidence (person, time, case and purpose, and authority); and,
- Any unavoidable changes to the digital evidence, the justification for the change, and the person responsible.

### **3.3 Evidence handling and analysis processes**

Section 4.3e of the SANS ISO/IEC 27035 document states that “Clear incident investigation procedures can help to ensure that data collection and handling are evidentially sound and legally admissible. These are important considerations if legal prosecution or disciplinary action might follow.” (p. 4). Section 8.2.5 further states that there should be the use of appropriate “IT based investigative techniques and tools, supported by documented procedures” (p. 37) and that the investigation should “should be conducted in a structured manner, and, as relevant, identify what may be used as evidence, whether for internal disciplinary procedures or legal actions” (p. 37).

The SANS ISO/IEC 27035 document (SABS, 2013) placed emphasis on the secure collection, storage, and preservation of digital evidence, and that this is continuously monitored in the event that the evidence is required for legal proceedings or disciplinary hearings. This is important as it indicates that there needs to be adequate rigour in the handling of digital evidence for disciplinary hearings. Therefore, those who are appointed to gather electronic evidence for disciplinary hearings, or are chairing the hearings, need to be adequately trained to ensure the fair and proper admissibility of electronic evidence during hearings. In particular, it needs to be shown that the “records are complete and have not been tampered with in any way; copies of electronic

evidence are provably identical to the originals, any IT system from which evidence has been gathered was operating correctly at the time the evidence was recorded” (SABS, 2014: 75).

### **3.4 Regulatory compliance of policies**

Annex E of the SANS ISO/IEC 27035 deals with legal issues regarding the organisational policies and procedures. Internal policies and procedures relevant to the disciplinary hearing should be verified to be compliant to legal regulatory requirements (SABS, 2013), and these policies need to be in place at the time of the alleged infringement.

## **4. Challenges of digital evidence in disciplinary hearings**

Due to the nature of digital evidence, there are a number of challenges that might arise. In particular, as disciplinary hearings are not necessarily as formal as court proceedings, there may be a more ‘relaxed’ approach to gathering digital evidence, resulting in poor procedure. This section will consider these challenges at various stages on the digital forensic cycle.

### **4.1 Pre-examination phase**

Many challenges may arise during the pre-examination phase that may affect the quality of evidence and analysis. These may include:

- Inadequately identifying potential sources of digital evidence; and,
- Illegally obtained evidence due to:
  - *Inadequate authorisation for collection;*
  - *Faked authorisation; and/or,*
  - *Misrepresentation for the evidence collection;*
- Poorly transported/transmitted/stored evidence, resulting in questions of its authenticity and accuracy;
- Evidence modification, due to:
  - *Unavoidable circumstances in the collection process;*
  - *Poor training of the investigators;*
- *Evidence tampering, by the target or accuser prior to collection, of by the person extracting/collecting the evidence;*
- Incomplete evidence and chain of custody, due to:
  - *Inadequate or improper documentation of the evidence and processes;*
  - *Incomplete evidence, such as a single email provided by a complainant out of context rather than an entire email trail;*
  - *System changes over time, e.g. the email server deleting emails the back-up process;*
- Fragmented processes between the forensic practitioner (outsourcing the work) and the practitioner acquiring the device or evidence.

### **4.2 Examination phase**

During the examination phase, the following challenges may arise:

- Inadequately defining or confirming the examination specifications (e.g. scope and time period under investigation);
- Inadequate validation of the evidence or tools;
- Modification of evidence during the analysis process;
- Improperly documented evidence from the pre-examination phase or during the examination phase;
- Analysis against irrelevant documents, such as policies and/or legislation that are outdated or had not come into affect at the time of the alleged infraction.

### **4.3 Post-examination phase and disciplinary hearing**

Challenges that can arise during the post-examination phase and/or the disciplinary hearing can include:

- Misrepresentation of the evidence and/or analysis in the report;
- Misrepresentation of the evidence and/or report during testimonies (by any party);
- Chair not understanding digital evidence or the challenges thereof;
- Excessive reliance on opinion as opposed to reality.

The last point is of particular interest due to the nature of legislation in South Africa. As digital evidence is often considered as 'hearsay', incorrect verbal evidence based on opinions may be given priority over digital evidence that reflects reality. For example, a system administrator stating that a system is secure does not change the fact that there could be a misconfiguration, which could be shown by digital evidence (in this case system logs). The very nature of a misconfiguration or security vulnerability is that it might be unknown; therefore multiple individuals may testify to the state of a system, but these are incorrect.

### **4.4 Other challenges**

It should also be noted that poor organisation processes and 'internal politics' may affect these processes. Due to circumstances within an organisation, there may be pressure to achieve a quick or specific outcome. In such cases, proper investigation may be completely bypassed and the case proceeds directly to a disciplinary hearing based on incomplete evidence by a complainant. In addition, a case may be erroneously (or intentionally) assigned to a department or individual that does not have the necessary skills. In addition, approval processes may be poorly defined and informally done, such as sending an email. Should the incident involve emails, and the integrity of the email server is under question, it is then inappropriate to use the body text of an email to request authority.

In addition to the hearing chair's understanding of digital evidence being a concern (as indicated in Section 4.3), the skills and education of digital forensic practitioners in South Africa are also shown to be below par (Jordaan and Bradshaw, 2015). This implies that there is the potential for poor quality evidence and analysis being presented to a chair without adequate experience, which can negatively affect the fairness of the disciplinary proceedings.

## **5. Recommendations**

Given the challenges of accepting and managing digital evidence, a number of recommendations are provided below to aid organisation, investigators employed by or contracted by the organisation, and chairs of disciplinary hearings, in ensuring proper process is followed and the proceedings are fair. This is important in order to protect organisations from litigation due to poor processes or techniques coming to light.

### **5.1 Documentation of processes and document templates**

Adequate and transparent organisational processes are important to aid in the collection of digital evidence, and successful disciplinary processes, otherwise questions regarding the legitimacy of the evidence and process could be raised. At its most basic level, there should be approved process documents for the handling of digital evidence in relation to disciplinary action, and be referenced by the relevant organisational policy(ies). The process document should be available to all employees, along with the organisational policies. The presence of the process document will help a chair in assessing the admissibility of evidence based on due process being followed.

In addition, organisations should have approved templates for chain of custody and evidence collection authorisations. An example template for chain of custody is shown in Figure 2. The authorisation for evidence collection should require explicit signing by the relevant person, with a second signature to prevent collusion. Copies should be filed by both the approvers and the requestor, so it is possible to verify the authorisation and its contents should the need arise. Under no circumstances should approvals for evidence collection be provided solely in the body of an email.



## 5.2 Collection of digital evidence

Not all organisations can afford to have a specialist digital forensic investigator permanently employed. They could then outsource this function, which may become expensive if there are many cases. Alternatively, individuals within the information technology department can receive basic training to enable them to securely extract the necessary information, such as email accounts, system logs, or other information.

- Digital evidence gained from the information technology department should highlight the following in the documentation:
  - Who assisted the forensic practitioner and the individual's designation level;
  - Who extracted the evidence;
  - The process that was used for extraction and the technology of tool that was used; and,
  - Who transferred the information and how the transfer of the information was conducted.

Figure 2: Example chain of custody template

A proposed process for collecting digital evidence is provided below. As emails are often a key feature in and may present a more complex situation due to verification, privacy and archiving concerns, the process will place emphasis on a case of obtaining emails as digital evidence. This process emanates from the authors' experience and guided by Buonocore (2021), Lazic and Bogdanoski (2018), and Mora (2009).

The forensic acquisition can be conducted either off-site or on the site of the client (e.g. in the site of the Information Technology section). Once the collection of emails and the preservation of the evidence has been implemented, the forensic practitioner needs to establish how the target's email folder was preserved. For example, where the email(s) reside and their state, and the complete set of emails needs to be confirmed. In addition, the investigator needs to establish if any of the emails have been modified, (either by the target or email administrator), even if the email is in a draft format; Mora (2009) discusses this for Microsoft Outlook servers. Where possible, there should be confirmation of the extracted contents, for example from server backups or email archiving solutions. During these processes, the metadata of the emails should also be confirmed. For example, the date and time stamps information is crucial to ensure the correct version of an email of interest is being analysed, as backup copies may differ or to check for modifications. In addition, the time and date stamps can be used for validating email evidence to the email folder records.

It is imperative that the forensic practitioner has an indication of the location of the evidence either saved to the target's local device(s) such as a laptop, tablet, smartphone or desktop, and establish if any synchronization is evident between the mail server and the device(s). In consultation with the email administrator, the email account should be migrated to backups, and the backup can then be copied. However, there also needs to be an awareness of offline storage, as the target may save or store emails off-line in a variety of file formats. The backups become important to confirm the emails of interest if extractions from the server are not possible. For completeness, the investigator should also request the email server logs.

At this point, an exact copy of the evidence can be copied, and a suitably qualified practitioner should verify this copy. Once this verification is received, only then can the evidence be associated to be an original. The investigator should also further verify the process followed to this point, the copying, and any analysis performed. The findings can then be documented and either a written report or affidavit is produced. A peer or senior should review the report or affidavit and as part of a quality assurance review; this is crucial as it provides confirmation of whether an independent forensic practitioner can arrive at the same findings and conclusions.

In some cases, it is prudent that the system administrators are also given an opportunity to outline the organisation's server and network architecture, server functions, data storage and security. For example, the management and security of the email accounts will be relevant to the above example process. It is important to note that due to the nature actual of technologies, supporting documentation should be provided to support the opinion of individuals when providing this information.

To summarise the process as illustrated by the above example:

- Extraction of data needs to be performed, often in conjunction with the assistance of the IT function;
- The extraction processes should be verified;
- Data should be verified with the aid of backups, system logs, and meta-data (particularly time and date stamps);
- The verified data can be considered to be original and used as evidence;
- The processes, including analysis and copying of the evidence, should be verified;
- The document reporting the findings can be produced, and should undergo a QA process.

### **5.3 Additional recommendations**

This section presents some additional recommendations that do not fit into the above recommendations, and are not extensive enough to warrant their own section. The fits recommendation is surrounding skills and training. Those staff members who are involved in conducting disciplinary hearings, such as chairs, senior line-managers, and HR, should be provided with training that explicitly covers the challenges and requirements of digital evidence. This will minimise failed disciplinary hearings and accusations of misconduct or unfairness in holding the disciplinary hearing. It is particularly important for any individual who is expected to chair a hearing to be fully aware of both the legal and technical challenges of digital evidence in order to make sound decisions in allowing evidence and in their decisions.

In addition, should non-specialists be used in the process of extracting or collecting evidence (e.g. system administrators in the IT function), that they are provided with elementary training in digital forensics to provide a degree of confidence in any actions they take as part of an investigation.

Organisations also need to consider the security of system logs in order to assure the integrity of the logs should they be required as evidence. To support this, Accorsi (2009a; 2009b) proposes secure protocols for system log transmission and storage.

From a legislative perspective, the relevant labour laws need to be revised in order to explicitly consider digital evidence, with particular guidance to the use of digital evidence in disciplinary hearings. This will aid in ensuring organisations are guided and adequately equipped to use digital evidence in a fair and acceptable manner.

Organisations need to begin considering privacy concerns in the investigation and hearing processes. Armknecht and Dewald (2015) propose technical discussion on protecting privacy in email forensics; however, given the

commencement of the Protection of Personal Information Act there should be a holistic approach to ensure compliance of the processes (both technical and business) in addition to introducing technical measures to aid in privacy preservation. Aligned to this, the monitoring, collection and archiving of communications should be compliant to RICA. This ensures that the relevant data will be available for use in internal investigation, and that it is collected and used in a legal and ethical manner.

## **6. Conclusion**

The use of digital communications and information is increasing, resulting in an increasing need to use digital evidence in disciplinary hearings. However, there are challenges in the use of such evidence and as disciplinary hearings are often not as formal as court proceedings, therefore it is haphazardly used in disciplinary hearings. Skills relating to digital evidence is important and digital forensics is evolving; therefore, forensic practitioners need to continuously upskill and keep abreast of new anti-forensic methods and media storages. In addition, chairs of disciplinary hearings need to be adequately trained to understand the challenges and volatility of digital evidence.

As people's jobs and livelihoods can depend on the evidence, organisations need to ensure a degree of rigour is provided to adequately use digital evidence in a fair manner. Preserving the integrity of the evidence is crucial, and presenting accurate and unbiased findings that have been verified is important. Organisations need to have clearly defined and documented processes for authorising the collection of evidence and documenting the processes and chain of custody. With the commencement of the privacy laws, organisations need to ensure they are compliant in using digital evidence for internal proceedings. In addition, South African labour laws need to be revised to improve guidance on digital evidence in disciplinary matters.

## **References**

- Accorsi, R., (2009a) Log Data as Digital Evidence: What Secure Logging Protocols Have to Offer? 33rd Annual IEEE International Computer Software and Applications Conference, Seattle, WA, pp. 398-403.
- Accorsi, R., (2009b) Safekeeping Digital Evidence with Secure Logging Protocols: State of the Art and Challenges, 2009 Fifth International Conference on IT Security Incident Management and IT Forensics, Stuttgart, Germany, pp. 94-110.
- Ali, K.M., (2012) Digital Forensics: Best Practices and Managerial Implications, 2012 Fourth International Conference on Computational Intelligence, Communication Systems and Networks, 196-199.
- Armknacht, F., and Dewald, A., (2015) Privacy-preserving email forensics, *Digital Investigation* 14, S127-S136.
- Buonocore, C., (2021), Computer Forensics in Today's World: Trends & Challenges, EC-Council, 5 February, [online], accessed 22 February 2021, <https://blog.eccouncil.org/computer-forensics-in-todays-world-trends-challenges/>
- Electronic Communications and Transactions Act, Act 25 of 2002, Republic of South Africa.
- ISACA, (2015) Overview of Digital Forensics, [online], accessed 11 January 2020, [https://www.isaca.org/bookstore/bookstore-wht\\_papers-digital/whpodf](https://www.isaca.org/bookstore/bookstore-wht_papers-digital/whpodf)
- Jordaan, J., and Bradshaw, K., (2015) The Current State of Digital Forensic Practitioners in South Africa, 2015 Information Security for South Africa (ISSA), Johannesburg, South Africa, pp. 1-9
- Jordaan, J., (2016) Understanding Digital Forensics - Standards, Processes, and the Law, ISACA South Africa Chapter 2016 Annual Conference, 29-30 August.
- Krotoski, M. (2011) Effectively Using Electronic Evidence Before and at Trial, *United States Attorneys' Bulletin* 59(6), 52-71.
- Labour Court of South Africa, Port Elizabeth., (2018) Case P 60/2018, South African Legal Information Institute, [online], accessed 8 April 2021, <http://www.saflii.org/za/cases/ZALCPE/2020/6.pdf>
- Law of Evidence Amendment Act, Act 45 of 1988, Republic of South Africa.
- Lazic, L., and Bogdanoski, M., (2018) E-Mail Forensics: Techniques and Tools for Forensic Investigation, The 10th International Conference on Business Information Security (BISEC-2018), 20th October, Belgrade, Serbia, 25-31.
- Michalsons (2008). Guide to the ECT Act in South Africa, 25 September, [online] accessed 23 February 2021, <https://www.michalsons.com/blog/guide-to-the-ect-act/81>.
- Michalsons (2021). POPI Regulations in South Africa Explained, 22 February, [online], accessed 8 April 2021, <https://www.michalsons.com/blog/pop-i-regulations-popia-regulations/12417>.
- Mora, R. (2009) Analysis of e-mail and appointment falsification on Microsoft Outlook/Exchange, SANS Institute, [online], accessed 22 February 2021, <https://www.sans.org/blog/analysis-of-e-mail-and-appointment-falsification-on-microsoft-outlook-exchange/>
- Mushtaque, K., Ahsan K., and Umer, A., (2015) Digital Forensic Investigation Models: an Evolution study, *Journal of Information Systems and Technology Management* 12(2), 233-244.
- Nortje, A. (2016) The Admissibility of Electronic Documents in Court Proceedings, *Polity*, 21 June, [online], accessed 8 April 2021, <https://www.polity.org.za/article/the-admissibility-of-electronic-documents-in-court-proceedings-2016-06-21>.
- Protection of Personal Information Act, Act 4 of 2013, Republic of South Africa.
- Regulation of Interception of Communications and Provision of Communication-related Information Act, Act 70 of 2002, Republic of South Africa.

***Trishana Ramluckan, Brett van Niekerk and Harold Patrick***

- Sabillon, R., Serra-Ruiz, J., Cavaller, V., and Can, J.J., (2017) Digital Forensic Analysis of Cybercrimes: Best Practices and Methodologies, *International Journal of Information Security and Privacy* 11(2), 25-37.
- Sonjica, N., (2021) Constitutional Court declares provisions of Rica unconstitutional, Times Live, 4 February, [online], accessed 8 April 2021, <https://www.timeslive.co.za/news/south-africa/2021-02-04-constitutional-court-declares-provisions-of-rica-unconstitutional/>.
- South African Bureau of Standards (2013) SANS 27035:2013 / ISO/IEC 27035:2011 Information technology — Security techniques — Information security incident management.
- South African Bureau of Standards (2014) SANS 27037:2014 / ISO/IEC 27037:2012 Information technology — Security techniques — Guidelines for identification, collection, acquisition, and preservation of digital evidence.
- Swales, L. (2018) An Analysis of the Regulatory Environment Governing Hearsay Electronic Evidence in South Africa: Suggestions for Reform – Part One, *Potchefstroom Electronic Law Journal*, 21, [online], accessed 8 April 2021, <http://www.saflii.org/za/journals/PER/2018/47.html>.

# Security and Safety of Unmanned Air Vehicles: An Overview

Sérgio Ramos, Tiago Cruz and Paulo Simões

University of Coimbra, CISUC, DEI, Portugal

[sdramos@dei.uc.pt](mailto:sdramos@dei.uc.pt)

[tjacruz@dei.uc.pt](mailto:tjacruz@dei.uc.pt)

[psimoes@dei.uc.pt](mailto:psimoes@dei.uc.pt)

DOI: 10.34190/EWS.21.027

**Abstract:** Cyber-physical systems permeate the fabric of our society, being a crucial part of what makes it possible. As such, ensuring their security is a primal concern that cannot be neglected, whether it relates to essential services, transportation or factories. But new scenarios and use cases are emerging, which require equal concern and care, as it is the case for Unmanned Air Vehicle (UAVs) technologies. Over the past years, several technological developments made it possible to create effective and inexpensive embedded computing and communications capabilities which, in their turn, were crucial for the development of modern UAVs. Such vehicles are quite diversified in terms of form and function, encompassing a wide range of implementations, from conventional drones to quad or octocopters. UAVs can be quite versatile, having found different applications, from leisure to warfare. Even though UAVs can help achieving better performance for transportation, surveillance, agriculture and healthcare, this technology shares most of the risks and dangers of IoT devices, since UAVs, or Drones, are devices with multiple sensors that can be integrated on Wireless Sensor Networks (WSNs), in a similar way as an IoT device. Drones are a doorway from the digital world to interact directly in the physical world. It is predicted that as more and more UAVs are being produced, passing from thousands to millions of drones on air every day, it increases the chance of someone hacking a drone and, for example, making it collide with a rotor of an airplane. Even though drones may not have sheer firepower, their speed and mass mean they still constitute a potential danger in many scenarios, with potentially catastrophic results. Therefore, it is vital to understand the security risks and vulnerabilities associated with UAVs in order to develop mechanisms to ensure proper safety requirements, also fostering general trust and paving the way for new services and opportunities. This paper presents an overview on the most recent developments on drone security and safety. It also presents a comparison between UAVs and IoT devices, highlighting the similarities between IoT and drone cybersecurity issues. Finally, it proposes possible solutions and countermeasures for other UAV vulnerabilities and their feasibility on the constrained field of drone ecosystems.

**Keywords:** unmanned air vehicles, internet of things (IoT), cybersecurity, cyber warfare

---

## 1. Introduction

Transportation is one of the main essential services for society. Since the first industrial revolution, it has accompanied the evolution and the development of technologies, moving goods, people, merchandise and commodities for longer distances and faster speeds. Particularly, and since the first days of human flight, air transportation has been accomplished mainly by aircrafts and helicopters manned by human pilots. However, in the last two decades the use of UAVs (more commonly known as Drones) have become commonplace, firstly for military purposes and then for civil applications.

Regardless of the type of aircraft, air transportation presents specific critical security risks: an error in midair may lead to the fall of an aircraft, resulting in material or even human losses. Drones are no exception for this concern, as any security or safety issues may cause serious incidents. Still, drones can be very versatile, having many applications in the civil domain, such as reconnaissance, photography, recreational usage and transportation. According to (SESAR JU, 2016), an economic impact analysis of the entire value chain for each of the areas of demand revealed the “potential for a European market exceeding EUR 10 billion annually by 2035”. Nevertheless, turning this forecast into reality will involve studying, researching and implementing proper UAV regulations, as well as adequate security and safety solutions.

This paper is organized as follows. In Section 2 we overview the main topics related with Unmanned Aerial Vehicles, followed by a brief discussion on communication technologies for UAV environments (Section 3). Next, we analyse the similarities between UAV and IoT domains (Section 4), followed by a discussion of cybersecurity aspects and recent developments for UAV ecosystems (Section 5). Section 6 concludes the paper.

## 2. Unmanned aerial vehicles

All UAVs are drones, but not all drones are UAVs, since the term refers to unmanned vehicles or systems, also including for instance unmanned submarines and unmanned land vehicles. This section provides an overview of the UAV ecosystem, with a specific focus on Unmanned Aircraft Systems (UAS).

The term UAV refers to a specific type of drones, which operate on air, while the term UAS refers to the whole system inherent to the UAV. For sake of simplicity, in this paper the terms UAV and Drone will refer to the same subject. Regarding operability, an UAV can be autonomous or remotely controlled, as in the case of Remotely Piloted Aircraft Systems (RPAS).

### 2.1 The UAS ecosystem

The UAS landscape encompasses a diversified set of concepts, equipment, technologies and procedures, as well as its own terminology. This section overviews these aspects, providing the knowledge baseline required to fully grasp the specific characteristics of the domain. Figure 1 depicts a simplified vision of the UAS ecosystem, including the following key components:

- The Ground Control Station (GCS), typically the physical infrastructure used to remotely control and monitor the UAV Operation, facilitating all the necessary information to the human operator or system.
- The Flight Controller is the main responsible for controlling the UAV.
- Regarding communications, the term “data link” is often used to refer to the wireless links used to transport the information flow between the drone and the GCS (Yaacoub, Noura, Salman, & Chehab, 2020). Besides wireless local network technologies (e.g., Wi-Fi), commercial and military drones often resort to cellular networks (such as 4G/5G) or satellite links for UAV control and communications.

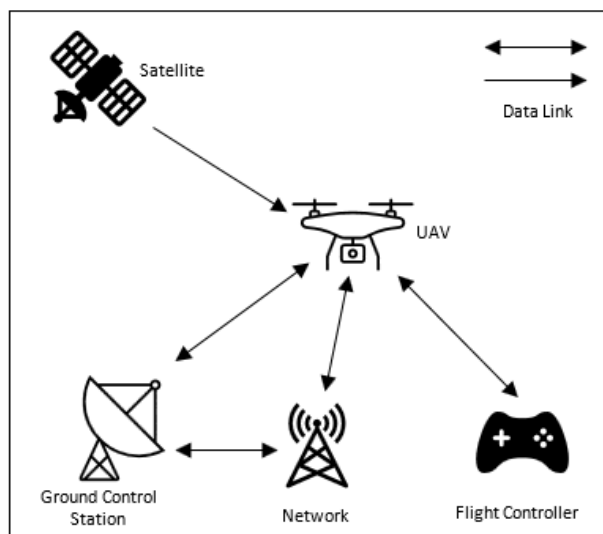


Figure 1: UAV ecosystem

### 2.2 Types of UAV

UAVs can be classified in three main types (Single-Rotor, Multiple-Rotor and Fixed-wing – see Figure 2), with a fourth category representing UAVs that combine more than one of them (Aircraft Compare Editorial Team, 2021).

Single-Rotor drones rely on only one main rotor to operate on air. An example of this type of drone would be a helicopter (even though it needs a second rotor to control its autorotation, it has a single the main rotor). Multi-Rotor Drones operate similarly, but rely on more than one rotor, typically four (quadcopters) or eight (octocopters). This type of drone is more common for civilian applications, being used for structure inspection, agriculture, photography and leisure, due to its maneuverability. Fixed-wing Drones may still have one or more rotors, but use fixed wings to fly, sharing the same principles of airplane aerodynamics. An example from this type of UAVs is the well-known Predator military drone (US Air Force, 2015).



**Figure 2:** Single-rotor drone, Quadcopter and fixed-wing Drone (Prodrone, 2021) (DJI, 2021) (US Air Force, 2015)

UAVs come in all sizes and shapes. Their weight may vary from a few grams (DJI, 2021) to several thousands of kilograms (U.S. Air Force, 2021), with the latter category being comparable to a small airplane in terms of size, with capacity for up to 8 or 10 people. These attributes are important for regulation purposes, in order to classify different UAV types and models.

Regardless of the type of UAV, there are two operation categories related to its distance from the operator: Visual Line of Sight (VLOS) and Beyond Visual Line of Sight (BVLOS). Additionally, the term VLL (Very Low Level) refers to a certain layer of the airspace where Visual Flight Rules (VFR) flights cannot operate (except for landing and take-off) (SESAR JU, 2019). While many Drones can operate within VLOS, only the more sophisticated models can operate in BVLOS, being able to cover longer distances.

Drones are complex systems, having different parts and specialized subsystems, where all the data is gathered and then sent to the remote pilot or system. Those components make drones very similar to Internet of Things (IoT) devices. For instance, a quadcopter drone typically has a main body part, standard and pusher propellers, landing gear, flight controller, RF unit, camera and sensors (e.g., gyroscopes, barometers, accelerometers, magnetometers, rangefinders, Inertial Measurement Units). Moreover, the information provided by the sensors may be augmented by means of artificial intelligence (AI) or sensor fusion algorithms, providing capabilities such as obstacle sensing and avoidance during the execution of the flight plan – a concept known as Detect and Avoid (DAA). Fully autonomous drones are also expected to gain more prevalence in the future, once proper Air Traffic Control (ATC) coordination mechanisms and regulatory frameworks are in place.

### 2.3 EU regulations and flight concepts

To the best of our knowledge, available literature provides little information about UAV Flight Concepts for civil aviation, which is hardly unsurprising considering the market maturity. Despite being seemingly more mature area, there is also a lack of information regarding military UAV operations, due to their sensitive nature.

As there are many types of drones, each with different purposes, there was the necessity to organize and classify them based on the type of missions and the associated risk levels. This makes it possible to predict and plan which resources and requirements to associate with each operation. The U-Space Concept, proposed by SESAR JU (SESAR JU, 2019), addresses this issue by defining three types of Flight Concepts: *open*, *specific* and *certified*.

The *open flight concept* category is not subject to any prior operational authorization, nor to an operational declaration by the UAS operator before the operation takes place (European Commission, 2019). According to (EASA, 2019), UAS operations do not belong to the *open* category when at least one of the general criteria listed in Article 4 of the UAS Regulation is not met (e.g., the drone does not carry dangerous goods and does not drop any material). Such operations usually fall into the *specific flight concept*, which requires operational authorizations, or under the *certified flight concept*, when more strict regulation and certification are mandatory (e.g. when the operation is conducted over assemblies of people or transporting people or dangerous goods with high risk for third parties in case of accident, in which case the Drone must specifically support such scenarios (European Commission, 2019)).

### 3. Communication technologies for UAV environment

Communications remain a crucial factor in UAV operations, regardless of their degree of autonomy. Whether used for direct operator flight control (with or without BVLOS), to provide transponder-like capabilities or to provide a command-and-control channel for strategic flight operations, communications play a vital role in enabling most operational scenarios, from simple 1:1 control to multi-UAV orchestration in ATC environments.

This section provides a quick overview of the UAV communications landscape, from scenarios to protocols, also briefly covering the radio technologies most commonly used for such purposes.

### **3.1 Types of communications**

As already mentioned, UAS heavily rely on different communication channels, not only to ensure control, but also to guarantee proper navigation and surveillance. The different communication scenarios in a UAS environment can be classified in four main types (Yaacoub, Noura, Salman, & Chehab, 2020):

- Drone to Drone – although not standardized, some drones come equipped with the capability of directly communicating with other drones.
- Drone to Ground Control Station – this is the most common type of communication, already standardized on several fields of application.
- Drone to Satellite – this type of communication uses GPS for controlling and monitoring the UAV from greater distances (e.g., in a BVLOS scenario). Satellite communications are deemed as reasonably secure and safe (Yaacoub, Noura, Salman, & Chehab, 2020), being heavily used in military applications, despite their high costs.
- Drone to Network – alternatively to Drone-to-Satellite communications or direct communications with the GCS, drones may communicate using cellular networks such as 5G.

### **3.2 Communication technologies and protocols**

Communications are a vital component of the UAS environment, being critical to its operation, security and safety. To support the different communication scenarios used UAS environments, there are several types of technologies adopted. For instance, the most commonly technologies currently used for communications with the GCS are Wi-Fi (IEEE 802.11), 4G and 5G. In many cases, such communications may be public, making them prone to eavesdropping or Man-in-The-Middle (MiTM) attacks. Drones may also form Mobile Ad-hoc Networks (MANETs), also called Flying Ad-hoc Networks (FANETs), possibly supported using Peer-to-Peer (P2P) technologies that may be vulnerable to Denial-Of-Service or Sybil attacks.

Furthermore, the L-band digital aeronautical communication system (LDACS), which is an upcoming air-to-ground digital communications standard for the aeronautic domain, might be leveraged for future data link applications. It operates in the L-band (around 1GHz), with excellent propagation characteristics with additional capabilities that allow it to be considered for future integrated CNS and Spectrum developments. Also, it has been demonstrated that LDACS is able to support Alternative Positioning Navigation and Timing (APNT), meaning it is a navigation backup for the Global Navigation Satellite System (GNSS). Furthermore, current studies are looking at how LDACS could support surveillance and RPAS command and control C2 link technologies and applications (Eurocontrol, 2021).

There is also a considerable diversity in terms of the UAS communications protocols being used for C2 and telemetry purposes. Some of those protocols were designed for generic use, such as the Message Queueing Telemetry Transport (MQTT) and the Data Distribution Service (DDS) protocols, while others were specifically designed for UAS (e.g., MAVLink). We will focus on these three examples, as they constitute a representative sample of the most popular approaches.

MQTT is a lightweight protocol that transports messages between devices, typically running over TCP/IP. It is an OASIS and ISO standard (ISO 20922 - MQTT, 2016), being designed for constrained and low-bandwidth, high-latency or unreliable IoT networks (MQTT.org, 2021). It also uses Transport Layer Security (TLS), requesting username/password and may optionally require certificates. This protocol is widely used for drone communications, for instance to control drones that transport medical samples (HiveMQ GmbH, 2021).

Figure 3 presents the publish/subscribe architecture of MQTT, including the following key components:

- The MQTT Client (subscriber and/or publisher) connects to the Broker and publishes and/or receives messages, by publishing or subscribing to specific topics, respectively.



- MQTT Broker – The Broker works as intermediary between devices, providing decoupling capabilities and ensuring that devices do not require a synchronous API to talk to each other, allowing better scalability compared to traditional client-server architectures. It persists the messages/topics published by the clients.

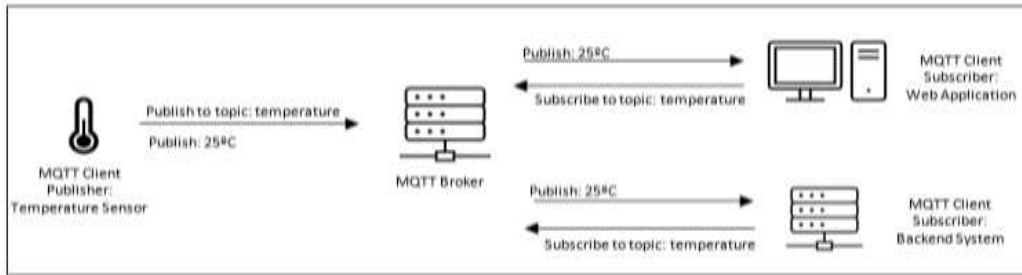


Figure 3: MQTT publish/subscribe architecture

DDS is a network middleware that uses a publish/subscribe pattern (DDS Foundation, 2021), allowing real-time and scalable data exchange. It is often used in ATC applications, medical devices, transportation systems and other applications that require real-time data exchange. Although similar to MQTT, it benefits from additional features, such as supporting more Quality of Service (QoS) parameters. While MQTT operates over a TCP connection (something that may pose an undesirable overhead in unstable connectivity scenarios), DDS resorts to a Real-Time Publish-Subscribe (RTPS) wire protocol for data transfer, which is transport-independent and may use UDP, TCP or other alternatives.

MAVLink is a lightweight messaging protocol for communicating with drones (Dronecode Project, Inc., 2021), following a hybrid publish/subscribe and point-to-point design pattern. The messages are defined in XML files defining the message set supported by a particular MAVLink system.

#### 4. Similarities between UAV and IoT

Among a wide range of applications, IoT devices can be found in smart vehicles, smart buildings, health applications, energy management, environmental management, food supply, production and industry lines management (Mosenia, 2017). This latter can be particularly useful to create decentralized industries, allowing to control, monitor and manage lines from a distance. Large industries are adopting IoT to acquire real-time information about their factories, thus evolving Industrial Control Systems (ICS) and SCADA systems. These industrial systems are also known as Industrial Internet of Things (IIoT). The ease in deploying IoT devices made those devices a top choice for most of aforementioned applications, revolutionizing entire industries. However, this fast growth and scalability also brought new challenges and security risks. Due to their pervasive and ubiquitous computing, adding the internet connection and the lack of security controls, made this type of devices prone to internet-based attacks, being used for instance to create large DDoS attacks (Antonakakis, et al., 2017). This section presents a brief overview of the IoT architecture and the UAS architecture, comparing the two models and pointing out similar security challenges.

##### 4.1 IoT reference architecture

IoT may require the connection of several different devices, which communicate with one another in a complex manner. For a better understanding of how these communication layers can be organized and studied, a simplified IoT Reference Architecture, composed of four different layers, is provided in Figure 4 and will be used in the rest of this paper.

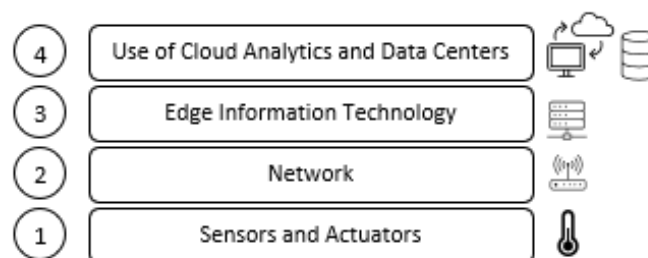


Figure 4: IoT architecture layers

Layer 1 relates with Sensors and Actuators (e.g., a pressure sensor or a temperature sensor). This layer interfaces with the physical world, being composed of devices capable of translating physical events into data and then transforming that data into information to be used by the applications of upper layers. It can also have actuators which react to changes in the environment. This layer is also known as the perception layer.

Layer 2 relates with the Network, where the data collected from the physical layers is transferred to the upper layers – using cellular communications, Wi-fi, Bluetooth Low Energy (BLE), RFID, NFC or other communication technologies. This layer also guarantees the communication between the devices from the first layer and between the components from this layer (Mosenia, 2017).

Layer 3 relates with Edge Information Technology and is also known as the Middleware Layer. It is responsible for the pre-processing and pre-analytics of the information transferred from the lower layers, which is then passed to the upper layers and systems. Since the amount of data can be very large (e.g., Big Data scenarios), without this extra layer the applications could be flooded with data, endangering their efficiency and performance. The processing of data is thus initiated on this layer (Mosenia, 2017).

Layer 4 relates with the usage of Cloud Analytics and Data Centres. The previously pre-processed information is processed in this last layer and presented to the users and the main applications for analysis. It is also known as the Application Layer or the Business Layer.

## 4.2 IoT security challenges

The attacks against IoT, presented in Table 1, can be analysed from four perspectives, according to (Francesca Meneghello, 2019).

**Table 1:** Functionality of attacks against IoT

Type of Attack	Description
Ignoring Functionality	In this type of attacks, the functionalities of the device are ignored, and only the capability to connect the device to the internet or to other nodes is exploited. This opens a door to the system and allows the attacker, for instance, to create botnets.
Reducing Functionality	By controlling and reducing the functionality of a device, the attacker can ask for ransom or reduce the effectiveness of the whole system, causing malfunctions, loss of data and functionality, or even permanent damage to physical resources.
Misusing Functionality	After exploiting a device, the attacker can use it for purposes other than those the device was originally designed for. For instance, an attacker can exploit a heat system or a sound system and cause discomfort to the users by turning up the heat or by playing annoying sounds.
Extending Functionality	In this type of attacks, for instance, an IP camera can be hacked to be used against the owners of a house, by controlling the camera and watching them without permission.

Next, the potential weaknesses of each layer are discussed. The devices from Layer 1 are deployed in the physical environment, which makes them prone to physical attacks. Moreover, especially when using wireless technologies, their communications are also prone to interception by malicious agents. Further, those devices have limited computation capabilities and limited battery power, making them more prone to cyberattacks. For instance, a malicious agent can add another node to the system, sending Malicious Data, causing a DoS attack by depriving other nodes from sleep mode and thus preventing them from saving energy (Shivangi Vashi, 2017). Hardware Trojans are possible to succeed in unsupervised nodes by inserting or substituting parts of the hardware of the devices. Also, side-channel attacks are possible, by analysing the power consumptions and other information about the device’s lifecycle. Other issues faced by this layer are mainly related with authentication, confidentiality and access control.

Layer 2 is composed of routers, access points and other wireless or wired devices, facing even more threats than traditional information systems networks, due to the openness characteristics of IoT, with lack of strong security mechanisms and weak protocols. Known attacks associated with this layer include Sybil Attacks, Sinkhole Attacks, Sleep Deprivation Attacks, DoS attacks, Malicious code Injection, MiTM attacks, routing attacks and injection of fraudulent packets. Layers 3 and 4 represent the higher-level components of the architecture. They

are prone to some of the problems faced by common information systems. Some of the security issues faced by these layers are malware, DoS attacks, Phishing, Spear-Phishing and sniffing attacks.

To prevent and mitigate these attacks, it is important to know which characteristics are relevant to guarantee adequate security. Besides the fundamental Confidentiality, Integrity and Availability properties (commonly known as the C-I-A Triad) there are other, often overlooked aspects, that may be also important, such as accountability and non-repudiation. Ultimately, the priority and importance of certain characteristics may depend on the specific nature of each device or use case. For instance, while in a wireless implanted medical device, integrity and availability take priority over confidentiality, in industrial systems it is more important to guarantee availability prior to confidentiality. Nevertheless, and regardless of the fact that security priorities may change according to specific use cases, markets or industries, implementation of proper defence-in-depth strategies may ultimately require paying attention to all the relevant aspects.

### 4.3 Comparison between UAVs and IoT

UAVs typically include a number of sensors, interconnected or connected to a central unit, that use a transmitter/receiver to communicate with the GCS or the flight controller. Those sensors are relatively close to one another, compartmentalized in the same space. In comparison, traditional IoT environments are much more open and may be much larger (see Figure 5), up to thousands of devices, and adding/removing devices is a less rigid process. Those devices can be interconnected to one another or communicate through routers or gateways. This way, while in traditional IoT it is easy to add or subtract a device, in UAVs the manufacturer designed and developed the aircraft based on a fixed (or at least very restricted) number of sensors.

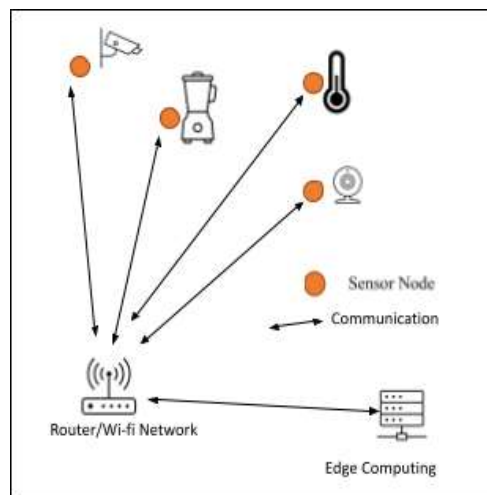


Figure 5: IoT architecture and sensors disposition

Figure 6 proposes a UAV architecture reference model. This reference model can be directly compared with the IoT reference architecture provided in Figure 6, in order to better highlight the similarities between these two domains. This UAV reference model also features four different layers. Layer 1 corresponds to UAV Sensors and Actuators, i.e., the physical layer where the UAV sensors will react to physical events, capturing data and transmitting that data to the actuators and/or to the upper layers of the architecture. Layer 2 corresponds to the Network and Data Link. This layer represents the communications and the data link between the UAV and the GCS. Layer 3 corresponds to the Ground Control Station, being responsible for acquiring and presenting the data sent by the UAV to the Drone Operator, as well as for sending commands to the UAV. Finally, Layer 4 corresponds to Cloud Analytics and Data Centres, being therefore related to the Business layer.

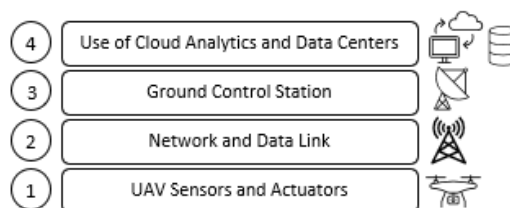
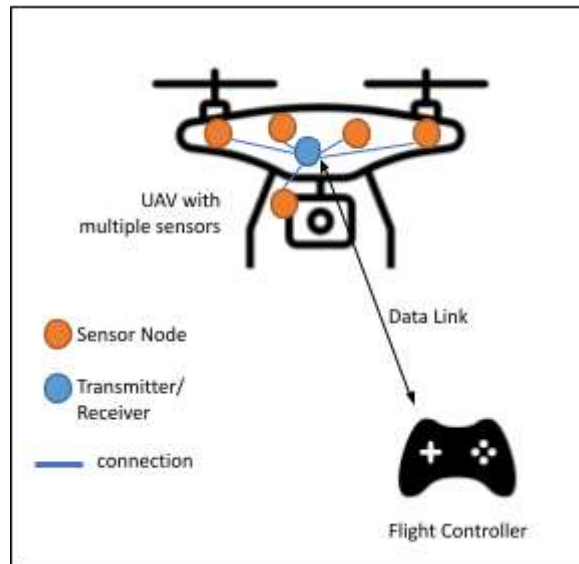


Figure 6: UAV architecture reference model

Despite the apparent similarities between the IoT and UAV reference models, they often differ when it comes to integration. IoT architectures usually encompass a diversified array of communications and control requirements, which are usually linked to the specific nature of a process or use case scenario. Thus, while some IoT systems may require real-time deterministic communications and control, others may be perfectly suited to rely on asynchronous communication mechanisms, supported by delay-tolerant networks or scenarios with unstable connectivity. For UAVs (which often incorporate a diversified array of sensor and actuator components, both for operational and mission-related purposes – see Figure 7), tracking and control requirements are not as strict as in typical hard-real time process loops, but there still are certain safety thresholds that must be ensured for reliable synchronous control – this is partially true even for (semi)autonomous UAVs, which have specific requirements for mission control and tracking purposes.



**Figure 7:** UAV architecture and sensor disposition

## 5. Cybersecurity and safety for UAV ecosystems

Drones have offered new ways of transport, surveillance and monitoring, being used in several areas and due to their versatility, both for military and civilian applications. However, those systems might present a set of vulnerabilities and threats, some of them shared with IoT devices, that need to be addressed. Regarding UAV Security and Safety, it's vital to ensure that the communication channels provide confidentiality, integrity, availability, authentication and non-repudiation properties (Yaacoub, Noura, Salman, & Chehab, 2020), although for communications the most valuable and prioritized characteristic is integrity, ensuring proper communication between air (UAV) and ground (ATC, UAV operator). Such requirements depend on a set of crucial capabilities, namely:

- Authorization, allowing to assign privileges to personnel in the UAS context and implement Role-Based Access Control (RBAC) mechanisms.
- Authentication, providing multi-factor authentication, mutual authentication mechanisms (especially relevant for Machine-to-Machine communications), use of Perfect Forward Secrecy (PFS) and robust encryption techniques.
- Auditing/Accounting, implementing proper mechanisms for log collection, aggregation and storage, supporting auditing procedures for post-mortem incident analysis or identification of non-conform procedures.

Regarding UAV Cybersecurity, Table 2 presents the most common cyber-attacks against drones and the correspondent security countermeasures that can be implemented.

**Table 2:** UAV cyberattacks and security measures

<b>Drone Cyber-Attacks</b>	<b>Security Measure</b>
Malware	Hybrid lightweight IDS, Control access, system integrity solutions and multi-factor authentication
Backdoor Access	Hybrid lightweight IDS, Vulnerability assessment, Multi-factor robust authentication scheme
Injection/modification	Machine-Learning hybrid IDS, Time stamps, Message authentication or Digital signature
Fabrication	Assigning privilege, Multi-factor authentication, Message authentication or Digital signature
Scanning	Hybrid lightweight IDS or Honeypot, Encrypted traffic/stream
Eavesdropping	Securing communication/traffic, secure connection
Man-in-the-Middle	Lightweight hybrid IDS, Multi-factor authentication & lightweight strong cryptographic authentication protocol
Skyjet	Lightweight IDS at the physical layer, Strong & periodic passwords, strong encryption algorithm, Security Awareness
Wi-fi Jamming	Redundancy, Frequency hopping, frequency range variation
De-authentication	Lightweight IDS, Security Awareness, Frequency hopping, frequency range variation
Replay	Frequency hopping, time stamps
Buffer overflow	Lightweight IDS, proper validation
ARP Cache Poison	Lightweight IDS, Encryption
Denial of Service	Redundancy
GPS Spoofing	Return-to-base, frequency range variation

Next, we provide a brief discussion of the most common attacks identified in Table 2:

- Injection/Modification – in this attack malicious or modified data is injected, leading to malfunctions on the drone or infections with malicious content.
- Fabrication - this attack aims at disrupting authenticity by trying to gain privileges to inject data or to access drone components.
- Skyjet – this is another Wi-Fi attack, presented by (He, Chan, & Guizani, 2017), in which an attacker searches for drones in the vicinity, takes control of those drones, turning them into zombie drones, and uses then the aircrack-ng tool to deauthenticate the drone and assume control over it.
- Wi-fi jamming – this type of attack jams all the wireless communications within a specified coverage area. Although this attack has some limitations, it can be catastrophic if succeeded during a long period of time, since UAS control may rely on this technology.
- Replay – a DoS attack in which data is intercepted and delayed or retransmitted at a later stage.
- Buffer Overflow – a DoS attack in which the network is flooded with requests to disrupt the drone connection.
- ARP Cache Poison – a Man-in-the-Middle attack which exploits the Address Resolution Protocol (ARP) messages, associating the attackers’ MAC Address to the target IP address, therefore allowing him to intercept the victim’s traffic (The MITRE Corporation, 2021). This technique can also be used to disconnect the Drone from the controller and/or the GCS.
- GPS Spoofing – probably one of the most known attacks against drones, GPS Spoofing refers to the type of attack which tries to give false GPS signals to the Drone receivers. This can cause malfunctions and even control over the UAS, for instance, as it may be headed to an unwanted place instead of the base station. Since civilian GPS systems are typically not encrypted, due to cost and strategy reasons, it is very difficult to counter this type of attacks (Yaacoub, Noura, Salman, & Chehab, 2020).

Besides the UAS ecosystem cyberthreats that may be exposed by certain exploitable vectors, there are other relevant issues to be accounted for, namely: technical, operational and natural issues. The first are a direct consequence of the complex nature of UAS as systems of systems, composed of many electronic components prone to fail, such as batteries, propellers, software and even communications. Operational issues may arise as a consequence of lack of operator skills or deficient flight planning, putting the UAS in danger with potential risks of collateral damage. Finally, natural issues such as wind and rain also constitute a serious threat for UAS

operation, as they may push components beyond their acceptable operation thresholds in terms of environmental parameters and mechanical stress.

Communications, Navigation and Surveillance (CNS) constitute the three fundamental pillars on top of which operational drone safety is built upon. Navigation helps leading the aircraft to the intended destination(s). Surveillance refers to monitoring its operating status and location (by simple visual contact or through more sophisticated techniques, such as GPS or radar). Communications guarantee that information is passed back and forth through the right channels and between the involved parts – a vital component of UAS security and safety. The main difference between manned and unmanned aviation is the physical location of the human controller: while on the former the pilot is onboard, on UAS the controller is located on a remote pilot station. This imposes new risks to be considered, during both the pre-flight and the in-flight phases, to protect these three fundamental services. Any successful attack during any of these phases may lead to catastrophic results, but threats may not present the same risk level for both phases. For instance, GPS jamming has more impact during in-flight, whereas in pre-flight (while the drone is on the ground) it represents a less serious threat.

To better understand this, it is important to note which services or procedures correspond to each flight phase.

The *pre-flight phase* comprises the pilot training, strategic conflict resolution, communication with the ATC, weather information, traffic information, drone maintenance and inspection. During this phase, some services can be targeted for cybersecurity attacks, for instance to acquire information about the drone mission, to compromise some of the components that may be critical during the flight phase, or even to provoke a Denial-of-Service situation. To achieve this, attackers may recur to well-known information systems threats, such as malware, backdoors, virus, trojans, MiTM, DoS and ARP cache poison. To help mitigate these risks, Ground Control Stations, UAS Operators and U-Space Service Providers (USSP) can rely on known existing standards and guidelines, such as the NIST Cybersecurity Framework for Critical Infrastructures (National Institute of Standards and Technology, 2018) or ISO 27002 (ISO 27002, 2013), to mention a few.

During the *in-flight phase* the UAV may need to communicate with the ATC, the Ground Control Station and the Flight Controller. This phase also encompasses weather information, traffic information, CNS monitoring and tactical conflict resolution. During in-flight, attackers may focus on more mission-critical components, such as the CNS infrastructure, the ATC communication or the drone communications. In this type of scenario, attackers may use techniques such as injection/modification, fabrication, scanning, eavesdropping, MiTM, skyjet, wi-fi jamming, replay, buffer overflow, ARP Cache Poison, DoS and GPS Spoofing.

To help addressing this variety of threats, threat-trees like those proposed by (Almulhem, 2020) can be very beneficial for designing cybersecurity mechanisms. This approach helps modeling and representing threats in a tree-shape diagram to a particular system and from a high-level perspective. Although it does not ensure complete representation of all possible threats, it offers an initial point of start. It is important to mention that the threat-tree may differ for different systems architectures, even within the UAS domain. To further analyze possible threats in a system and ones that might have not been addressed using the previous technique, simulated environments (Mairaja, Babab, & Javaid, 2019) may help researching for vulnerabilities and security issues in drones and communication protocols, as well to better understand UAV technologies, reducing experimentations time and costs.

The specific nature of the UAS cybersecurity domain, as well as its intersection with the particularly strict requirements of the aeronautical industry, can only be adequately tackled by means of a significant investment in two crucial areas.

The first area is Regulation and Standardization. Contrary to the IoT industry, the aviation industry is a well-regulated and standardized sector, benefiting from synergies derived from common efforts and oriented goals. Since the early 2000's, Eurocontrol and other aviation entities, like the International Civil Aviation Organization (ICAO), are regulating, publishing standards and guidelines for the UAS industry, that help protecting the different stakeholders. Additionally, this industry benefits from adopting standards from related areas, such as the NIST Cybersecurity framework or the ISO 27002, serving as guidelines for the implementation of Security and Information Management systems.

The second area is Research & Development. As part of a unified industry approach, there is a considerable effort, both in the EU and internationally, to research, develop and deploy a standardized Air Traffic Management (ATM) infrastructure. This effort has spawned several developments, such as the U-Space Traffic Management initiative (UTM) (SESAR JU, 2019). An example of this common effort is the development of a common Public Key Infrastructure framework for the System Wide Information Management (SWIM), led by Eurocontrol and involving a large number of industry partners. This framework is planned to be deployed until the end of 2021, ensuring interoperability of digital certificates within Europe and other regions and offering security and encryption for the communications (Eurocontrol, 2021).

Overall, these developments show to which extent the industry is committed towards the development of proper regulations and support infrastructure, which are deemed vital for the mass introduction of UAV-supported services expected to have a disruptive effect in many different sectors, such as healthcare, delivery of goods and surveillance or emergency relief support.

## **6. Conclusion and future work**

This paper presented the main topics related to UAS security, considering the whole UAV ecosystem and regulations, and pointing out different communication technologies and protocols relevant for the security of UAS communications. The similarities between UAV and IoT were explored in order to provide a different perspective of UAV cybersecurity, including the highlighting of several attacks and corresponding countermeasures and the discussion of recent developments in the aviation industry.

The implementation of standards and the application of risk and cybersecurity frameworks are important and should be used to make the UAV ecosystems more resilient to cybersecurity attacks. Furthermore, research of new vulnerabilities and the development of new architectures for UAV ecosystems, using simulated environments, plays a key role to deploy faster and better solutions.

## **Acknowledgements**

This work was partially funded by the European Union in the scope of the BUBBLES Project (SESAR JU, 2020), funded in the scope of the SESAR Joint Undertaking (SESAR JU), under the Horizon 2020 Research and Innovation Program (agreement number 893206).

## **References**

- Aircraft Compare Editorial Team. (2021). *Types of Drones*. Retrieved from <https://www.aircraftcompare.com/blog/types-of-drones/>
- Almulhem, A. (2020). Threat modeling of a multi-UAV system. *Transportation Research Part A: Policy and Practice*, 142, 290-295. doi:10.1016/j.tra.2020.11.004
- Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., . . . Zhou, Y. (2017). *26th USENIX Security Symposium*. Retrieved 2021, from Usenix.org: <https://www.usenix.org/system/files/conference/usenixsecurity17/sec17-antonakakis.pdf>
- DDS Foundation. (2021). *DDS Foundation*. Retrieved from <https://www.dds-foundation.org/what-is-dds-3/>
- DJI. (2021). *DJI Mini 2*. Retrieved 2021, from <https://www.dji.com/pt/mini-2?site=brandsite&from=nav>
- DJI. (2021). *DJI Phantom*. Retrieved 2021, from <https://www.dji.com/pt/phantom>
- Dronecode Project, Inc. (2021). *Mavlink*. Retrieved from <https://mavlink.io/en/>
- EASA. (2019). *Acceptable Means of Compliance and Guidance Materials*. Retrieved 2021, from <https://www.easa.europa.eu/sites/default/files/dfu/AMC%20%26%20GM%20to%20Commission%20Implementing%20Regulation%20%28EU%29%202019-947%20%E2%80%94%20Issue%201.pdf>
- Eurocontrol. (2021). *LDACS*. Retrieved from <https://www.eurocontrol.int/system/l-band-digital-aeronautical-communication-system>
- Eurocontrol. (2021). *SWIM Common PKI*. Retrieved 2021, from <https://www.eurocontrol.int/project/swim-common-pki-and-policies-and-procedures-establishing-trust-framework>
- European Commission. (2019). *Regulation on unmanned aircraft systems and on third-country operators of unmanned aircraft systems*. Retrieved 2021, from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R0945&from=EN>
- European Commission. (2019). *Rules and procedures for the operation of unmanned aircraft*. Retrieved 2021, from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R0947&from=EN>
- Francesca Meneghello, M. C. (2019). IoT: Internet of Threats? A Survey of Practical Security Vulnerabilities in Real IoT Devices. *IEEE Internet of Things Journal*, 8182-8201. doi:10.1109/JIOT.2019.2935189
- He, D., Chan, S., & Guizani, M. (2017). Drone-Assisted Public Safety Networks: The Security Aspect. *IEEE Communications Magazine*, 218-224. doi:10.1109/MCOM.2017.1600799CM

- HiveMQ GmbH. (2021). *Matternet*. Retrieved 2021, from <https://www.hivemq.com/case-studies/matternet/>
- ISO 20922 - MQTT. (2016). Retrieved 2021, from ISO: <https://www.iso.org/standard/69466.html>
- ISO 27002. (2013). Retrieved from ISO: <https://www.iso.org/standard/54533.html>
- Mairaja, A., Babab, A. I., & Javaid, A. Y. (2019). Application specific drone simulators: Recent advances and challenges. *Simulation Modelling Practice and Theory*, 94, 100-117. doi:10.1016/j.simpat.2019.01.004
- Mosenia, A. (2017). *Addressing Security and Privacy, Challenges in Internet of Things*. ArXiv.
- MQTT.org. (2021). *MQTT.org*. Retrieved 2021, from <https://mqtt.org/faq/>
- National Institute of Standards and Technology. (2018). *Framework for Improving Critical Infrastructure Cybersecurity*. Retrieved from <https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf>
- Prodrone. (2021). *Prodrone PDH-GS 120*. Retrieved 2021, from <https://www.prodrone.com/products/pdh-gs120/>
- SESAR JU. (2016). *European Drones Outlook Study*. Retrieved 2021, from [https://www.sesarju.eu/sites/default/files/documents/reports/European Drones Outlook Study 2016.pdf](https://www.sesarju.eu/sites/default/files/documents/reports/European_Drones_Outlook_Study_2016.pdf)
- SESAR JU. (2019). *CORUS ConOps Vol2*. Retrieved from <https://www.sesarju.eu/sites/default/files/documents/u-space/CORUS%20ConOps%20vol2.pdf>
- SESAR JU. (2020). *Bubbles project*. Retrieved from <https://bubbles-project.eu/>
- Shivangi Vashi, J. R. (2017). Internet of Things (IoT): A Vision, Architectural Elements, and Security Issues. *Proceedings of the International Conference on IoT in Social, Mobile, Analytics and Cloud, I-SMAC* (pp. 492-496). Palladam, India : Institute of Electrical and Electronics Engineers Inc. doi:10.1109/I-SMAC.2017.8058399
- The MITRE Corporation. (2021). *ARP Cache Poisoning*. Retrieved 2021, from <https://attack.mitre.org/techniques/T1557/002/>
- U.S. Air Force. (2021). *Global Hawk*. Retrieved 2021, from <https://www.af.mil/About-Us/Fact-Sheets/Display/Article/104516/rq-4-global-hawk/>
- US Air Force. (2015). *MQ-1B Predator*. Retrieved 2021, from <https://www.af.mil/About-Us/Fact-Sheets/Display/Article/104469/mq-1b-predator/>
- Yaacoub, J.-P., Noura, H., Salman, O., & Chehab, A. (2020). Security analysis of drones systems: Attacks, limitations, and recommendations. *Elsevier - Internet of Things*, 11. doi:10.1016/j.iot.2020.100218



# The Rising Power of Cyber Proxies

Janine Schmoldt

University of Erfurt, Faculty of Economics, Law and Social Sciences, Germany

[janine.schmoldt@uni-erfurt.de](mailto:janine.schmoldt@uni-erfurt.de)

DOI: 10.34190/EWS.21.068

**Abstract:** More and more states support cyber proxies. States work with proxies rather than sanctioning them. This is because cyber proxies are incredibly useful – they not only enhance the cyber warfare capabilities of the supporting states, they also provide them with a degree of plausible deniability. This ascent a future superpower status. China for instance uses cyber proxies in order to “deter the United States (...)” which “may ensure eventual strategic parity with the United States in technological and military prowess” (Hjortdal 2011). Simultaneously, cyber proxies are seen as an inherent threat to the stability of nation states as they are capable of subverting the national and political stability. Is then the support of cyber proxies not inconsistent with the aim to limit potential threats to the stability of nation states? Through the support of nation states, cyber proxies can enhance their technical skills with the result, that they elevate their political role. Cyber proxies have become extremely powerful, supporting them means that they are directed “away from operating against the state” (Hang 2014). Thus, even more governments have decided that it is better to work with rather than against cyber proxies. But this also leads to the question of whether states can be held responsible under international law for the actions of the cyber proxies and hackers.

**Keywords:** cyber proxies, cyberwarfare, patriotic hackers, international law

---

## 1. Introduction

Technology now controls our daily life to an unprecedented level from water supplies, communication, electricity generation and almost every aspect of our western developed culture “making it increasingly susceptible to computer network attacks and other cyber operations during armed conflicts” (Harrison Dinniss 2013). Those computer network attacks have already been launched by patriotic hackers. One of the earliest instances of patriotic hacking can be found in the Kosovo war when pro-Serbian (or anti-Western) hackers, such as the so-called *Black Hand*, conducted cyber-attacks against NATO, UK and US computers with the goal of disrupting their military operations by deleting all data (Geers 2017). But the Black Hand was not the only hacker group that launched cyber-attacks during the Kosovo conflict. Chinese patriotic hackers also conducted cyber-attacks after a U.S. B-2 stealth bomber dropped five precision-guided munitions on the Chinese Embassy in Belgrade (Wu 2007). Because of the hackers’ involvement in the conflict, “[t]he war in Kosovo has intensified as hackers on either side of the conflict try to take over or block Web servers around the world” (Messmer 1999). The hackers not only launch cyber-attacks on their own initiative, they were also used as “proxies to mount cyber-attacks by nation states [such as North Korea or Iran] (...)” (Dahan 2013). Tim Maurer also makes clear: “To project cyber power, (...) states rely on hackers that do not wear uniforms and are not part of the intelligence community – cyber mercenaries or, more badly, cyber proxies” (Maurer 2018). Yet, the United Nations General Assembly Document A/68/98\* of 2013, emphasizes that “States must not use proxies to commit internationally wrongful acts. States should seek to ensure that their territories are not used by non-State actors for unlawful use of ICTs” (UN General Assembly 2013). Although 15 nation states, including China, the United States, France and Russia agreed on this, cyber proxies are still used. Nation states support proxies rather than sanctioning them. And even more governments have decided that it is better to work with hackers and cyber proxies although their power inevitably grows. China for instance, inevitably made clear that their cyber proxies and patriotic are “a powerful threat to the Chinese government because they are a large politically active organization capable of subverting Chinese political stability” (Hang 2014). This clearly shows that the use of cyber proxies introduces risks such as the principal-agent problem: “The further detached an agent from a principal’s control, the higher the risk that the agent carries out an action that the principal did not intend” (Maurer 2018). Because of that, several questions arise: (1) Why do states (still) use cyber proxies? (2) Are the states aware of the risks of using cyber proxies? (3) Are states responsible for the actions of their cyber proxies under international law? The aim of this article is to answer these questions in order to shed light on the current role of cyber proxies. But before discussing the questions, the first part will present a definition of cyber proxies and give a brief literature overview. Afterwards, the second part will then elaborate why states use cyber proxies. The third part then describes how the power of cyber proxies rise through their state-relationship before a last part will deal with the question of whether states can be held responsible for the actions of cyber proxies under international law.

## 2. Literature review and definitions

It is meanwhile a common belief that hackers constitute a new type of actor in International Relations and International Politics. Particularly in political conflicts, there is often talk about accompanying cyber-attacks conducted by hacker groups where it is unclear how their activities link to the state. Erica Borghard and Shawn Lonergan recently illustrate that states “often form relationships with cyber proxies when they lack an independent ability to conduct cyber operations and/or seek plausibly to deny involvement in a cyber operation” (Borghard, Lonergan 2016). Thereby, states, governments and the hackers themselves “mutually benefit from establishing a relationship, although the specific nature of the goods and services being exchanged varies” (Borghard, Lonergan 2016). Governments could get political and material goods provided by cyber proxies. Through cyber proxies, states can “achieve objectives indirectly, thus shielding governments from political stakeholders who may shy away from direct action” (Borghard, Lonergan 2016). Cyber proxies also offer “political benefits to governments who may want to deny their involvement in a particular operation or even domain” (Borghard, Lonergan 2016). Thus, they provide governments with political goods. In terms of material goods, cyber proxies can offer the technical ability to launch cyber-attacks, “including the manpower” (Borghard, Lonergan 2016). Tim Maurer states that “most states’ proxy relationships fall into one of three main types: *delegation, orchestration, and sanctioning*” (Maurer 2018). While the delegation concept describes “proxies on a tight leash with the state and under the state’s effective control” (Maurer 2018), orchestration means that proxies receive equipment, funding, or other tools “but no specific instructions” (Maurer 2018). A “shared ideological bond” (Maurer 2018) is crucial for the concept of orchestration. Sanctioning “builds on the counterterrorism literature’s concept of passive support” (Maurer 2018). This topic also found its way into international law debates. Michael Schmitt and Liis Vihul as well as the Tallinn Manual on the International Law Applicable to Cyber Operation for instance dealt with questions of attribution in cyberspace and when cyber activities of hackers and cyber proxies may “be attributed to a state as a matter of international law” (Schmitt, Vihul 2014). Yet, this contribution will also use an international law approach in order to analyse the state responsibility for the actions of their cyber proxies. But before addressing the relationship of states and cyber proxies, it is, however, important to understand what is meant by cyber proxies. By tending towards Erica Borghard’s and Shawn Lonergan’s definition, this contribution regards cyber proxies as non-state hackers who conduct cyber-attacks or cyber operations in order to achieve “a political objective on behalf of a patron state” (Borghard, Lonergan 2016). It is also important to point out that cyber proxies “can be either individuals or loosely organized units comprised of patriotic hackers, hacktivists or cyber terrorists” (Borghard, Lonergan 2016). Yet, this contribution will focus on patriotic hackers as cyber proxies since they have a greater impact than hacktivists or cyber terrorists. In contrast to hacktivists, patriotic hackers not only engage in civil disobedience, they also get active in actual armed conflicts. Driven by patriotism, patriotic hackers hack with the aim to defend their homeland or national pride whereas hacktivists are essentially activists who are driven by the defence of a political or social issue that does not touch upon their home country or residence. Having said this, we can now focus on the state-proxy relationship.

## 3. The benefits states gain from cyber proxies

More and more states use cyber proxies because cyber proxies are incredibly useful – they not only enhance the cyber warfare capabilities of the supporting states, they also provide them with a degree of plausible deniability. Cyber proxies are able to avoid attribution as they can launch cyber operations “with little or no attribution” (Applegate 2011). Scott Applegate illustrates that

*“[t]hese types of attacks can be carried out inexpensively, with little or no political ramifications to the nation-state, and give the attacker a distinct asymmetric advantage. If a nationstate can covertly initiate, fund, or guide such attacks, (...) [relying on cyber proxies] they can potentially achieve their political objectives without the burden of attribution or the need to adhere to the Law of Armed Conflict” (Applegate 2011).*

Therewith, cyber proxies protect governments from the legal consequence of conducting cyber-attacks (Applegate 2011). Similarly, they also shift the problem of attribution. The problem is no longer ‘who carried out the attack’ but rather ‘does the state have/had control and command over the cyber proxies’. The cyber-attacks against Estonia illustrate this. Indeed, Russian patriotic hacker Konstantin Goloskokov has claimed responsibility for the cyber operations. However, the attacks “were in part attributable to the Nashi youth activist group, but it is unclear whether the Russian Federation had a hand in the group’s operations” (Schmitt, Vihul 2014). Related to this, Michael Schmitt and Liis Vihul point out that

*“the relatively high levels of support that are required before a state can be held responsible for the activities of non-state groups or individuals, as distinct from their own responsibility for being involved, creates a normative safe zone for them” (Schmitt, Vihul 2014).*

Although cyber proxies provide plausible deniability, not all states want that as the 2001 cyber-attacks against the U.S. illustrate. In 2001, following the collision of a US EP-3 reconnaissance aircraft and a Chinese F-8 PRC fighter cyber-attack were related to the U.S. (Wu 2007). During this event, the Chinese government was “not concerned with denying its responsibility for cyber-attacks” (Hang 2014). On the contrary, “the Chinese government encouraged computer-savvy citizens to deface American websites to express their displeasure” (Hang 2014). Similarly, in 1999, when Taiwanese President Li Teng-Hui advocated his two-state-theory China again “launched the Patriotic Hackers and encouraged other hackers to join during the next crisis with Taiwan” (Hang 2014). Nevertheless, cyber proxies are used to maintain plausible deniability. Next to that, they also enhance the cyber warfare capabilities of states.

Another reason why states work with cyber proxies is to increase their cyber warfare capabilities. This “can ensure its ascent to a future superpower status” (Hjortdal 2011). Cyber proxies are especially relevant for states “when they lack the capability (...) to conduct offensive cyber attacks” (Borghard, Lonergan 2016). North Korea for instance “is incentivized to work with cyber proxies because it lacks indigenous capabilities to conduct independent offensive operations” (Borghard, Lonergan 2016). Furthermore, states use cyber proxies in order “to respond to an external threat (i.e., a state adversary)” (Borghard, Lonergan 2016) or to an internal one. In many cases, cyber warfare capabilities constitute an external threat. The United States for instance threatens China through their cyber warfare capabilities and therewith, they constitute an external threat for China. Because of that, China uses cyber proxies in order to “ensure eventual strategic parity with the United States in technological and military prowess” (Hjortdal 2011). However, states also use cyber proxies “to co-opt potential threats to the regime, employing domestic groups that may pose a risk to regime stability to operate against adversaries, thus enabling the government to monitor those groups and keep them occupied” (Borghard, Lonergan 2016). This directly leads to the next reason why states use cyber proxies.

Through the support of cyber proxies states have a certain degree of control through which they can direct cyber proxies away from launching cyber-attacks against the state. Cyber proxies such as patriotic hackers constitute an internal threat especially to China. The cyber-attacks against Indonesia illustrate this. In 1998, many ethnic Chinese living in Indonesia were killed and raped. This constitutes a big dilemma to the Chinese government. On the one hand, the Chinese government could not accept the assaults towards the Chinese ethnicity. But on the other hand, their “long term foreign policy stance has been not to interfere with other countries internal affairs, and not to be meddled in by foreign powers” (Wu 2007). Wu highlights: “If not for the Internet, the Chinese government might have been able to turn a deaf ear toward it and hope it gradually died down” (Wu 2007). However, Chinese patriotic hackers took “justice into their own hands” (Wu 2007) and launched cyber-attacks against Indonesia. Already in 1998 after the Indonesian cyber-attacks, patriotic hackers were able to make Indonesia believe that China declared war. During a press conference on 9 August 1998, the Indonesian government made clear:

*“We hope Chinese people can keep calm. The past events are our internal affairs. Today, Chinese hackers are coming and we are confused. They are interfering in our problems. Chinese government not only interferes with our internal affairs, but also incites hackers to attack our network. We are very disappointed (...). Certainly, it is hard to believe this thirty-hour long persistent online attack was not an organized operation. If not for the absence of the air defense alarm, I would have believed that China had declared war against us” (Wu 2007).*

As opposed to Indonesia assumption, the Chinese government was not involved in the cyber-attacks against Indonesia. On the contrary, Chinese patriotic hackers launched the attacks on their own initiative by forming the *Chinese Hacker Emergency Conference Centre* in order to “set a course of action against Indonesia” (Denning 2011, Henderson 2007, Wu 2007). Twenty years ago, patriotic hackers were able to alter China’s politics. They forced the Chinese government to react. Wu emphasizes that patriotic hackers have the potential to even “transform China’s political structure” (Wu 2007). Cyber proxies are “now far better situated to influence political outcomes, even within the international area” (Hang 2014). Therewith, they constitute an internal threat to nation states. Cyber proxies are powerful actors, supporting them means that they are directed “away from operating against the state” (Hang 2014). Because of that domestic political control, China for instance, supports patriotic hackers: “The People’s Republic of China supports Patriotic Hacker attacks to control Patriotic

Hackers through fostering nationalism among hackers” (Hang 2014). Next to China also other states such as North Korea or Iran work with cyber proxies. And “even more authoritarian governments have decided that it is better to work with rather than against Patriotic Hackers” (Hang 2014). Through this support, cyber proxies will become more powerful (Hang 2014) because they also benefit from the states.

#### 4. The benefits cyber proxies gain from states

Indeed, cyber proxies also benefit from the support they receive from states. Borghard and Lonergan illustrate that “cyber proxies receive asymmetrically greater benefits from governments than the reverse” (Borghard, Lonergan 2016). Through the support of nation states, cyber proxies are able to finance themselves. Furthermore, they can enhance their technical skills through valuable equipment, including soft- and hardware. The fact, that states work with cyber proxies rather than sanction them shows clearly that states somehow also offer legal protection. In exchange for launching cyber-attacks and/or operations, states can provide money or employment. We must be keenly aware of

*“the magnitude of the financial incentive for cyber proxies, both in terms of direct transfers of money to groups from the state, and the indirect, but more lucrative, benefit to groups of being able to continue to conduct illegal activities in cyber space” (Borghard, Lonergan 2016).*

States can also provide valuable hard- and software “that may be very expensive or over which they may have a monopoly” (Borghard, Lonergan 2016). With this equipment, cyber proxies can exploit and test their capabilities or even develop other malware and viruses. Except equipment, states “can grant cyber proxies access to the full range of the Internet itself – states can remove restrictions, such as unfettered access and increased bandwidth, enabling the platform to function on an advantageous way for proxies” (Borghard, Lonergan 2016). Furthermore, states can also help to train other cyber proxies, they can “transfer technical knowhow” and even “provide information that can be used to exploit existing vulnerabilities” (Borghard, Lonergan 2016). Therewith, cyber proxies and hackers elevate their political role and they get a voice within the political arena. Operation Allied Force illustrates that cyber proxies can force states to react in the way the proxies want them to. Hackers already have the power to direct states and to influence a political outcome. During Operation Allied Force, the NATO air war against Yugoslavia during the Kosovo War in 1999, the Chinese embassy in Belgrade became accidentally the target of a bombing. As a result, Chinese patriotic hackers conducted cyber-attacks against the U.S. Therewith, “[t]he war in Kosovo has intensified as hackers on either side of the conflict try to take over or block Web servers around the world (...)” (Messmer 1999). This “cyber self-defense war” (Wu 2007) was also a factor why “Washington agreed to pay \$28 million for property damage, \$4.5 million to the families of the killed and injured, and in April 2000 fired one CIA operations officer involved in the targeting” (Lampton 2002). But the NATO campaign also illustrates another fact: Cyber Proxies or patriotic hackers get active in armed conflicts in order to support their homeland. The Russo-Georgian war of 2008 similarly illustrates that patriotic hackers are involved by conducting cyber-attacks against Georgia in order to defend Russia (Tikk, Kaska, Vihul 2010). In the same year, patriotic hackers were involved in *Operation Cast Lead* and in Operation Pillar of Cloud during the Gaza War between Palestinians and Israel where they stole lists with contact information of soldiers and reservists and warned them in joining the war (Dahan 2013). The Second Lebanon war, the tensions between Israel and Iran as well as the current conflict in Syria also illustrate that patriotic hackers are involved in armed conflicts. This shows clearly that in theory, international conflicts no longer require a state or state sanction. Cyber proxies from different states fight or hack against each other without governmental direction. The hacker war following the collision of the US EP-3 reconnaissance aircraft and a Chinese F-8 PRC fighter as well the cyber-attacks against Indonesia illustrate this. Cyber proxies can also threaten the internal affairs of a state and therewith the stability of states. All in all, the state’s assistance will enhance the proxies technical and “elevate these hackers into a more prominent political role” (Hang 2014). Rather than sanctioning the illegal activities, states “offer safe haven to cyber proxy groups, both in granting physical security within a state’s sovereign territory, as well as legal protections” (Borghard, Lonergan 2016). Almost all cyber proxies operate outside of the domestic (and international) legal rules and are therefore, susceptible to prosecution and jail. By working with cyber proxies, states are aware of the fact that they refrain from enforcing existing law.

#### 5. States, proxies and international law

States are aware of the activities of the proxies and tolerate them “in spite of having the capacity to do otherwise” (Maurer 2018). In these situations, the states are “clearly turning a blind eye to the proxies’ offensive actions” (Maurer 2018). This however, leads to the question of whether states can be held responsible for the

actions of the cyber proxies under international law. The International Law Commission's Articles on Responsibility of States for Internationally Wrongful Acts make clear:

*"The conduct of a person or group of persons shall be considered an act of a State under international law if the person or group of persons is in fact acting on the instructions of, or under the direction or control of, that State in carrying out the conduct" (International Law Commission 2001, Article 8).*

Thus, the actions of cyber proxies are attributable to the state if the hacker is acting under the instruction or direction or control of a State. The degree of control that must be exercised was a key issue for the International Court of Justice (hereafter ICJ) in a *Case Concerning the Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v. United States of America)*. The question was

*"whether or not the relationship of the contras to the United States Government was so much one of dependence on the one side and control on the other that it would be right to equate the contras, for legal purposes, with an organ of the United States Government, or as acting on behalf of that Government" (International Court of Justice 1986, para.109).*

The Court applied what has later been called the *effective control test*. A State can only be held responsible for the acts of non-state actors if the State "directed or enforced" (International Court of Justice 1986, para. 115) the non-state actor. The ICJ thus

*"Decides that the United States of America, by training, arming, equipping, financing and supplying the contra forces or otherwise encouraging, supporting and aiding military and paramilitary activities in and against Nicaragua, has acted, against the Republic of Nicaragua, in breach of its obligation under customary international law not to intervene in the affairs of another State" (International Court of Justice 1986, para.3).*

The Appeals Chamber of the International Criminal Tribunal for the Former Yugoslavia has also addressed the degree of control over non-state actors by making clear:

*"One should distinguish the situation of individuals acting on behalf of a State without specific instructions, from that of individuals making up an organised and hierarchically structured group, (...) Consequently, for the attribution to a State of acts of these groups it is sufficient to require that the group as a whole be under the overall control of the State" (International Criminal Tribunal for the Former Yugoslavia 1999, para. 120).*

Specific instructions are not required for the actions of individuals, as opposed to the *effective control test*, where specific instructions or directions concerning the performance of an action are required in order to attribute the action to a state. The Chamber thus identified another degree of control, namely the *overall control test*:

*"In order to attribute the acts of a military or paramilitary group to a State, it must be proved that the State wields overall control over the group, not only by equipping and financing the group, but also by coordinating or helping in the general planning of its military activity. Only then can the State be held internationally accountable for any misconduct of the group" (International Criminal Tribunal for the Former Yugoslavia 1999, para. 131).*

In the case of cyber proxies, it seems clear, that states have effective and/or overall control over the proxies, especially through the relationship of delegation, orchestration, and sanctioning. States can be held responsible for the actions of a cyber proxies under international law if they have effective and/or overall control or more precisely, if the cyber proxy is acting on the instructions, direction, or control of that state.

## **6. Conclusion**

More and more states work with cyber proxies due to three facts. First, cyber proxies avoid attribution and provide plausible deniability. Therewith, they protect states from the legal consequence that (illegal) cyber operations may cause. Second, cyber proxies enhance the cyber warfare capabilities of states when they lack offensive cyber capabilities. Third, cyber proxies co-opt threats as states gain a certain degree of control over the hacker through which they can direct cyber proxies away from launching cyber-attacks against the state. On the other hand, cyber proxies receive greater benefits from states which lead to the fact that their power will inevitably grow. By gaining a state's support, cyber proxies can finance themselves and enhance their technical skills. Therewith, cyber proxies can not only develop new tools they can also train other hackers. Furthermore,

states offer legal protection with the result, that hackers will not be prosecuted for illegal cyber operations. However, states who work with cyber proxies are responsible for the actions and operations of the proxies, if the proxies are acting on the instructions, directions, or overall and/or effective control of that state. All in all, cyber proxies get “a seat at the table of political change” (Hang 2014). Cyber proxies can already threaten the internal affairs of a state and therewith the stability of states as past cyberwars have shown.

## References

- Applegate, S.D. (2011) Cybermilitias and Political Hackers – Use of Irregular Forces and Cyberwarfare, *IEEE Security and Privacy*, Vol 9, No. 5, pp. 16-22.
- Borghard, E. and Lonergan, S. (2016) “Can States Calculate the Risks of Using Cyber Proxies?”, *Orbis: a quarterly journal of world affairs*, Vol60, No. 3, pp. 395-416.
- Dahan, M. (2013) *Hacking for the homeland: Patriotic Hackers Versus Hacktivists*, in: (editor) Hart, D. (2013) *The Proceedings of the 8th International Conference on Information Warfare and Security*, Published by Academic Conferences and Publishing International Limited.
- Denning, D. (2011) *Cyber Conflict as an Emergent Social Phenomenon*, in: (editors) Hold, T.J. and Schell, B.H. (2010) *Corporate Hacking and Technology-Driven Crime: Social Dynamics and Implications*, Information Science Reference.
- Geers, K. (2017) “Cyberspace and the Changing Nature of Warfare”, [online], NATO Publications, <https://www.sto.nato.int/publications/.../SMP-IST-076-KN.pdf>.
- Hang, R. (2014) “Freedom for Authoritarianism: Patriotic Hackers and Chinese Nationalism”, [online], The Yale Review of International Studies, <http://yris.yira.org/essays/1447>.
- Harrison Dinniss, H. (2013) *Participants in Conflict – Cyber warriors, patriotic hackers and the laws of war*, in: (editor) Dan, S. (2013) *International Humanitarian Law and the Changing Technology of War*, Martinus Nijhoff Publishers.
- Henderson, S. (2007) *The Dark Visitor – Inside the World of Chinese Hackers*, Published by Scott Henderson.
- Hjortdal, M. (2011) “China's Use of Cyber Warfare: Espionage Meets Strategic Deterrence”, *Journal of Strategic Security*, Summer, Vol 4, No. 2, pp. 1-24.
- International Court of Justice. (1986) Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America), Merits, Judgment, I.C.J. Reports (1986).
- International Criminal Tribunal for the Former Yugoslavia. (1999) Prosecutor v Dusko Tadic (Appeal Judgement), IT- 94-1-A, 15 July 1999.
- International Law Commission. (2001) Draft Articles on Responsibility of States for Internationally Wrongful Acts, Supplement No. 10 (A/56/10), chp.IV.E.1.
- Lampton, D.M. (2002) *Same Bed, Different Dreams: Managing U.S.- China Relations, 1989-2000*, University of California Press, Los Angeles.
- Maurer, T. (2018) *Cyber Mercenaries: The State, Hackers, and Power*, Cambridge University Press.
- Messmer, E. (1999) “Kosovo cyber-war intensifies: Chinese hackers targeting U.S. sites”, [online] CNN, <http://edition.cnn.com/TECH/computing/9905/12/cyberwar.idg/>.
- Schmitt, M.N. and Vihul, L. (2014) “Proxy Wars in Cyber Space: The Evolving International Law of Attribution”, *Fletcher Security Review* Vol1, No. 2, pp. 55-73.
- Tikk, E. and Kaska, K. and Vihul, L. (2010) *International Cyber Incidents: Legal Considerations*, Cooperative Cyber Defence of Excellence, Tallinn.
- UN General Assembly. (2013) A/68/98\* Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, 24 June 2013.
- Wu, X. (2007) *Chinese Cyber Nationalism: Evolution, Characteristics, and Implications*, Lexington Books, Plymouth.

# Connected, Continual Conflict: Towards a Cybernetic Model of Warfare

Keith Scott

De Montfort University, Leicester, UK

[jklsconfig@dmu.ac.uk](mailto:jklsconfig@dmu.ac.uk)

DOI: 10.34190/EWS.21.046

**Abstract:** "Our enemies are innovative and resourceful, and so are we. They never stop thinking about new ways to harm our country and our people, and neither do we." (George W. Bush) The purpose of this paper is to argue that to see 'cyber warfare' as a discrete form of combat, or as merely a combination of Electronic and Information Warfare, is a fundamental error. We must see 'cyber' as shorthand for 'cyberNETIC', and cyber warfare as a form of conflict which operates across all domains, and where action in one domain inevitably influences other zones of conflict. The UK military is seeking to reshape itself according to the concept of Integrated Operating, and this paper contends that such a model is essential. Marshall McLuhan defined World War 3 as 'a guerrilla information war with no division between military and civilian participation'; a cybernetic conflict is infinitely more complex, erasing the boundaries between kinetic and non-kinetic warfare, between civilian and military, and indeed between peace and war themselves. The paper will consider a scenario demonstrating what such a multi-domain conflict might be like, considering the use of non-human combatants operating in cooperation and against human forces, and the impossibility of maintaining a clear division between 'war' and 'operations other than war'. Ultimately, it will contend that the current structures of military forces are too rigid and rooted in earlier eras of warfare to allow us to respond effectively to the conflicts that await us in the all-too-near future. Norbert Wiener sought to avoid applying his knowledge of cybernetics to the military domain; this paper argues that it must be done. It is, in short, the most useful theoretical framework for waging hybrid, non-linear warfare.

**Keywords:** cyber warfare, strategy, hybrid warfare

---

## 1. Putting the 'cyber' back in 'cyberwarfare'

*"Use the word 'cybernetics', Norbert, because nobody knows what it means. This will always put you at an advantage in arguments." (attributed to Claude Shannon in letter to Norbert Wiener, ca. 1940s)*

The purpose of this assembly is to discuss Cyber Warfare and Cyber Security; the purpose of this paper is to argue for a redefinition of these terms, arguing that the word 'Cyber' has become so loose and imprecise in meaning that its use blinds us to a more helpful and productive model for understanding and mitigating the threats posed by new forms of conflict.

What do we mean by 'Cyber Warfare?' There are a multiplicity of possible definitions, which while all covering roughly the same domain, display shades of nuance specific to the concerns of the group or author coining them. Consider the following, which define 'Cyberwarfare' as:

- a. interstate use of technological force within computer networks in which information is stored, shared or communicated online (Green 1981)
- b. an extension of policy by actions taken in cyberspace by state actors (or by non-state actors with significant state direction or support) that constitute a serious threat to another state's security, or an action of the same nature taken in response to a serious threat to a state's security (actual or perceived). (Shakarian 2013)
- c. warfare grounded on certain uses of ICTs within an offensive or defensive military strategy endorsed by a state and aiming at the immediate disruption or control of the enemy's resources, and which is waged within the informational environment, with agents and targets ranging both on the physical and non-physical domains and whose level of violence may vary upon circumstances (Taddeo 2012).
- d. actions by a nation-state to penetrate another nation's computers or networks for the purposes of causing damage or disruption (Clarke 2010)
- e. conducting, and preparing to conduct, military operations according to information-related principles (Arquilla and Ronfeldt 1997)).

The differences are striking; for some, this is a form of conflict which focuses on the disruption/destruction of a state's informational assets through the application of non-kinetic, IT and AI-based tools (essentially a form of

Electronic Warfare), while for others, it implies a much broader set of actions, where non-kinetic action generates effects (which may cause damage equivalent or identical to kinetic force) far beyond the informational realm. Furthermore, the actors in such conflict may be nation states or non-state actors, and the targets may be military and/or civilian, attacking critical national infrastructure, financial institutions, and the beliefs of the citizenry (Information Warfare). In short, it is a form of conflict where lots of things are done to lots of people in lots of different ways by lots of different actors; as a theoretical framework for analysis and action, this is less than helpful. The one common feature the definitions have is an insistence that this form of conflict relies on the use of IT- and/or AI-enabled attacks; in other words, 'cyber' resides firmly within the Fifth Domain or Fifth Dimension of warfare, the informational realm (Fogleman 1995). Such a model is, I would argue, so imprecise as to be useless; it neglects the key nature of modern conflict (and of the modern world in general) – connection. The belated move to Multi-Domain Integration in military thinking (see Section 2.0) marks a recognition of the fact that dividing the world into 5 discrete domains is both illusory and dangerous; we cannot see the domains as separate, and crucially, we must not assume that the informational realm can be confronted in isolation. For the purposes of this paper, then, 'an act of cyber warfare' should be taken to mean 'any hostile action by a state or non-state equivalent actor launched against another state or non-state equivalent actor, which aims to cause kinetic and/or non-kinetic damage to the opponent's human/material/intangible assets, and which employs IT-based tools, networks and/or assets.'

'Cyber' has become a catchall term for 'anything involving computers'; this is both too vague and too precise; *computers* are not the important element, rather the system of integrated elements which make up the informational network which underpins and penetrates the modern world. Some of these elements are electronic, some are mechanical, some physical, some abstract (software does not 'exist' as a concrete artefact) – and many of them are human. The theoretical bedrock of this paper, and of the argument it advances, is that 'cyber' is meaningless; the term it derives from is what is important. We must turn, or return, to a *cybernetic* framework. The reason for this is simple; cybernetics offers a system-based model for understanding the informational network(s) within which Cyber Warfare operates, and it focuses on interconnection, and the interreactions within these networks of non-human *and* human elements.

'Cybernetics' may seem just as vague a term as 'Cyberwarfare', but that is to neglect an essential feature of the discipline; it is less a model for understanding a particular system (electronic, human, mechanical...) than a way of thinking about systems in general. The term was first used by Ampère (1834) to refer to political (i.e. human) governance, while the subtitle of Wiener's seminal text *Cybernetics* (1948) discusses 'control and communication in the animal *and* the machine' (my emphasis). By its very nature, cybernetics is interdisciplinary, and offers a model for analysing the interplay and interconnection of the various components of any system; it is, as Beer (1973) puts it, 'the science of effective organization'. As such, the adoption of a 'cybernetic mindset' as the basis for the consideration of contemporary and future conflict seems not only advisable but essential. As will be argued, both in terms of the complexity and protean nature of warfare in the Information Age, and in the growing convergence of human and non-human actors in the battlespace (human-machine teaming, autonomous weapons systems and augmentation of combat troops all falling within this ambit), the holistic, systems-based worldview offered by cybernetics seems to be unavoidable if we are to stand any chance of overcoming the threats we face. As with any discipline, cybernetic military thinking can follow many different paths. On the one hand, we have the more traditional C2C-based ideas of informational superiority seen in Cebrowski and Garstka's work on 'network-centric warfare' (1998); on the other, the work of Antoine Bousquet (2008a and b), whose theories of 'chaoplex warfare' discuss the challenges of waging a war which is non-linear, dynamic, and self-organizing, and where the opposing forces reject traditional notions of hierarchy and chains of command. It is this latter perspective which seems to offer the best approach to what we see emerging on the contemporary battlespace, as the following section will argue. In many of today's conflicts, there is no single 'enemy', rather a swarm of different groups and actors, sharing knowledge and resources, forming temporary alliances, and highly dynamic, ever-shifting formations. The cybernetic perspective, focusing on the network rather than individual agents/nodes within that network, is better equipped to respond to the current and future challenges. When conflicts operate across domains, collapse boundaries between state and non-state, private and public sector, military and civilian and kinetic and non-kinetic attacks, the *Gestalt* perspective of cybernetics offers a way of cutting through the complexity to understand the situation in its totality.



## 2. War everywhere, all the time: Cybernetic conflict

*this kind of war means that all means will be in readiness, that information will be omnipresent, and the battlefield will be everywhere. (Qiao and Wang 2020, Part 1).*

The starting point for grasping the nature of a truly cybernetic warfare is to recognise that 'warfare' itself manifests in many forms, both throughout history and in the present. Contemporary conflict can employ 'traditional' engagements between large forces of troops and matériel (e.g. the Battle of Medina Ridge during Operation Desert Storm), the use of non-kinetic IT-based 'ordnance' (such as STUXNET), the deployment of UAVS for reconnaissance, intelligence gathering, and targeted strikes, *and* still have troops fighting hand-to-hand 'with iron bars, rocks and fists' (Safi, Ellis-Petersen, and Davidson 2020). The theoretical division between 'conventional' and 'unconventional' war is, and always has been, illusory. As McFate (2019a) says, '[t]here is no such thing as conventional versus unconventional war – there is just war' (pp. 27-8). The norm is that there *is* no norm, or that rather, that a new paradigm is in play. It is undeniably true that what has been seen as 'cyberwarfare' exists, and has an ever-growing role to play in the battlespace, but this is only one facet of the wider cybernetic model. Previous military theorists, from Sun Tzu through von Clausewitz to Galula and Kitson and beyond, have each focused on one particular martial model; the future of warfare seems likely to privilege those who can think beyond traditional boundaries. The nature of cybernetic warfare is that it is uncertain, unbounded, and unrestricted. As Wiener (1980) said, in discussing the desire to divide neat boundaries between disciplines, '[t]here is no Maginot Line of the brain' (p.122.) It is striking that Wiener should make reference to one of the greatest examples of an outmoded military technology; the most interesting work on modern and future conflict comes from those who take a similarly flexible approach. In a key text for understanding this new vision of warfare, Qiao and Wang (2020) dismiss the concept of 'cyber' as a discrete domain:

*[information technology] is a synthesis of other technologies, so that it is part of them, and they are part of it, and this is precisely the most fundamental characteristic of the age of technological integration and globalization. (p.5)*

There will be those who read the preceding and argue that it is merely nothing more than rebranding as 'cybernetic warfare' the ideas that have been dubbed the 'Gerasimov doctrine' (Gerasimov 2016). This is incorrect, for the following reasons:

- a. Gerasimov did not create a truly new 'doctrine', but simply sought to 'develop an operational concept for Russia's confrontation with the West in support of the actual doctrine that has guided Russian policy for over two decades: the Primakov doctrine' (Rumer 2019);
- b. the idea that Gerasimov's ideas amount to a revolutionary new model of military thinking has been called into question, not least by one of the writers who initially helped to publicise it in the West (Galeotti 2016 and 2019);
- c. above all, while the 'doctrine' does suggest the use of both 'conventional' and 'unconventional' methods, and of combining soft and hard power, it does not advance a holistic model for using combined techniques simultaneously; it is not a truly multi-domain model.

I return to the earlier description of cybernetic warfare as 'uncertain, unbounded, and unrestricted'; these are the key aspects, as outlined below:

a. **UNCERTAIN:** one of the hallmarks of cyber-actions is that attribution for an attack is extremely difficult; online identities can be spoofed or concealed. The online domain offers potential for *maskirovka* which would be impossible in non-virtual battlespace. As is shown by the fact that the standard term for military deception is Russian, it is no surprise that Russia has been quick to develop informational deception as a tool of IW (Moore 2019). However, it is to be expected that the use of such tactics will become much more widely-used, as a means of low-risk offensive operations, whether as Information Warfare ('fake news' is not going to go away) or for attacks on financial markets and critical national infrastructure targets, as well as within a conventional military context. Not only will we be uncertain as to who the attacker is, it is more than likely that we will be unaware that an attack is actually occurring (consider the insertion of malware as part of apparently routine traffic).

b. **UNBOUNDED:** we should see cybernetic warfare as inextricably linked to the porosity and weakness of traditional boundaries; '[i]n a world where all things are interdependent, the significance of boundaries is merely relative' (Qiao and Wang 2020, p.189). If we examine the West, we can see a collapsing of the barriers between private and public sectors, between military and civilian realms (consider the ever-growing use of private military

contractors (McFate 2019b) between ‘conventional’ and ‘unconventional’ warfare, and, of utmost concern for those concerned with cybernetics, between human and non-human actors. Future conflicts will continue this process of collapse, and we should consider the implications of the following:

- i. **MULTI-DIMENSIONAL:** The recent UK Ministry of Defence Joint Concept Note on Multi-Domain Integration marks a step change in British military thinking; it does not simply argue that a modern fighting force must be configured to operate in all five domains, but that it should adopt as a fundamental principle the employment of actions which operate across domains. This is of course not an inherently radical concept (consider aerial bombardment of a land target), but the true conceptual shift comes with the embedding of the informational domain within all other domains; we might almost consider that there are not five domains, but an informational domain from which depend and within which operates activity in all the other domains. An informational attack on, for example, a navigational satellite affects not just that asset, but all other networked assets that depend on it; this could lead to the denial of communications, inaccurate targeting data for both human and non-human offensive forces, the disabling of fly by wire systems, and a range of other catastrophic results in land, sea and air forces simultaneously.
  - ii. **HUMAN-MACHINE INTERACTION AND CONVERGENCE:** Cybernetics, as stated earlier, deals with ‘animal and machine’; the modern world in general rests on a network (or network of networks) of human and non-human elements and actors, and the martial realm is no exception. Discussion of the interaction/convergence of human and machine in a military context inevitably leads to fears of two distinct nightmare scenarios. Firstly, machine intelligences achieve full autonomy; see, for example, *Slaughterbots*. This pseudo-documentary seems all-too plausible, given what is known about ongoing research into autonomous drone swarms (Marks 2020; Safi 2019). Secondly, there is the fear of human augmentation, of combat troops being ‘upgraded’ and made less human (oddly, cardiac pacemakers, stents, and hip replacements do not raise the same fears), as with the psychopathic half-machine assassin in Warren Ellis’ *Global Frequency* (2003). Again, the imagined future is an extrapolation of current research into augmentation technology (Emanuel et al. 2019; Lin 2012; Royal Society 2012); this can in turn be traced back to one seminal paper, Clynes and Kline’s (1960) article which coined the word ‘cyborg’, derived from ‘cybernetic organism’.
- c. **UNRESTRICTED:** ‘Unrestricted’ does not mean ‘unrestrained’; as defined by Qiao and Wang (2020), it signifies a holistic view of conflict, where any part of an enemy’s systems (military, political, economic, social...) is a legitimate target, and an ‘attack’ need not be kinetic:

*a single man-made stock-market crash, a single computer virus invasion, or a single rumour or scandal that results in a fluctuation in the enemy country’s exchange rates or exposes the leaders of an enemy country on the Internet, all can be included in the ranks of new-concept weapons. (p. 13)*

If applied as Qiao and Wang suggest, this form of ‘warfare’ is a truly integrated, network-wide, ‘supra-domain’ (their term) form of conflict, and it cannot be left solely to the military. A cybernetic warfare is multi- and trans-disciplinary, fought across all areas of human activity, in an integrated, interlinked chain of action and result; for this to succeed, it must be devised and governed across traditional managerial and administrative lines:

*only if we break through the various kinds of boundaries in the models of our line of thought, take the various domains which are so completely affected by warfare and turn them into playing cards deftly shuffled in our skilled hands, and thus use beyond-limits strategy and tactics to combine all the resource is of war, can there be the possibility that we will be confident of victory (Qiao and Wang 2020: 200-01).*

In the war that is coming, silos will be for missiles, not thinking.

### 3. Confronting cybernetic conflict

What, then, might a truly cybernetic attack look like? And how might it be combatted? In answer to the first question, we cannot yet say in truth; if we consider the classic triad of criminal investigation – motive, means, opportunity – then up until now there have been no shortage of motives for an attack, nor opportunities to launch one (and the multiple overlapping networks of systems of systems, vulnerable SCADA systems, and the Internet of Things have vastly increased the threat surface for on- and offline actions). What has been lacking are the means, or rather the ability to coordinate a range of complex and overlapping actions in real time; the use of AI-driven tools for planning, delivering, and coordinating a range of events across the domains,

automatically generating commands to the actors and transmitting misinformation to the targets will be a significant threat in the future.

There have been a plethora of fictional depictions of future war<sup>1</sup>, but perhaps the finest initial depiction of what a cybernetic attack might look like is presented in Philip Palmer's 2012 play *Red and Blue: Terror*. A British brigadier and Bradley Shoreham, a former colonel (who now runs corporate and government-level wargames) discuss various hypothetical ways of destroying London. Shoreham outlines a scenario which combines attacks that have previously occurred with as-yet unreal events:

*Mini-nuke in Central London – triggers all the protocols. We evacuate. Add sarin in the Tube. Throw in a series of bomb threats on the M25, M1, South Circular, North Circular, and you stymie all attempts at orderly evacuation. Mix in Mumbai-style terrorists in the suburbs; mass panic. Mobs in the street; crush injuries. Rioters throughout London, taking advantage of the sudden and total absence of a police force. A coordinated attack on many fronts. Everything happens at once.*

[...]

*West End theatre. A crowded place. Or a cinema, flooded with VX gas. Or leave an anthrax vial in Leicester Square – doesn't need to be broken, the rumour is enough. You leave the anthrax, then you tweet about it. Panic inevitably ensues. Or inject a rat with botulinum toxin and throw it into the underground system. Set fires in assorted London suburbs; there's nothing terrifies people so much as a good blaze on the 10 o'clock news. A sniper; shoot a few shoppers on Oxford Street with a military-issue sniper's rifle, and suddenly the streets will be empty. It's all about maximising fear.*

The key phrase here is 'Everything happens at once'; kinetic and non-kinetic events perturb all sectors of the social and informational networks simultaneously. The death toll and fear is merely one aspect of the attack; it generates shockwaves which create:

*the perfect economic, social, political and military storm. A terror attack; the threat of a mini-nuclear bomb in London. [the country at a standstill. The City destabilised. British bonds plummeting in value. Economic meltdown, political meltdown. A Third World War played out in the stock markets. The end of capitalism...*

Are we able to combat such an attack? The inability to coordinate an effective response to a single event (the current pandemic) or to combat the informational assaults of Fake News attacks, suggests not. Just as the adoption of a cybernetic approach to warfare requires the adoption of new supra-domain command and control structures, so the effective response to such actions will require root and branch revision of our current security structures. History, sadly, tells us that while such change can occur, it is only after the catastrophe has occurred. There needs to be much closer cooperation between domestic security, intelligence, and military authorities; instead of setting up the much-vaunted 'Space Force', the United States would be better occupied establishing an integrated cross-service multi-domain defence force, which can sit within existing structures and pivot into action whenever needed. This of course poses the question of how they will know when and where they are needed, given the number of possible attack points and the amount of data that will require monitoring.

In the Christmas edition of *The Spectator* last year, Dominic Cummings (the former key adviser to the UK Prime Minister) wrote about Stanislav Petrov, the Russian Officer who refused to notify his superiors that monitoring systems indicated an American nuclear strike was underway. Through his decision, Petrov prevented World War Three. Cummings argues that 'far greater intellectual and material resources ought to be deployed on such apparently low-probability, high-impact events' (Cummings 2020a). Cummings is an avowed evangelist for AI-driven analysis of Big Data as an analytical and predictive tool, and has argued that we are neglecting the benefits of working at the intersection of;

- the selection, education and training of people for high performance
- the frontiers of the science of prediction
- data science, AI and cognitive technologies (e.g Seeing Rooms, 'authoring tools designed for arguing from evidence', Tetlock/IARPA prediction tournaments that could easily be extended to consider 'clusters' of issues around themes like Brexit to improve policy and project management)

---

<sup>1</sup> See *inter alia*: Lewis 2018, Singer and Cole 2015, Krepinevich 2010, and the texts listed in Newman and Unsworth 1984.

- communication (e.g Cialdini)
- decision-making institutions at the apex of government.

(Cummings 2020b).

- in short, a technocratic solution to a cybernetic problem. Such an approach will almost inevitably fail; as McFate (2019a) says, of the Pentagon's Project Maven (an attempt to develop AI-led 'algorithmic warfare'), it will create:

*a case study in modern war ethics and the violation of privacy rights. It would take the guesswork out of the future by sucking in every email, camera feed, broadcast signal, data transmission – everything from everywhere – to know what the world is doing, with the omniscience of a god. The ancient Greeks had a word for this: hubris (p.50).*

The technological superiority of the US Army did not help them defeat the Viet Cong; nor have UAV strikes eradicated Al-Qaeda or ISIS/ISIL. The conflict in Northern Ireland may offer a better example of the approach to take; The IRA were not 'defeated' as such, rather they were outflanked by intelligence gathering, policing, military action, political negotiation and economic investment; a blended approach operating on all fronts simultaneously, covertly and overtly. McFate argues that:

*War is armed politics, and seeking a technical solution to a political problem is folly. Ultimately, brainpower is superior to firepower, and we should invest in people, not platforms. (p.57).*

- he is right, but he is only *partially* right. It should be remembered that cybernetics is concerned with the 'animal and the machine'; only a blended approach can truly hope to combat the supra-domain threats we face. Consider in closing an example taken from law enforcement; automated facial recognition is, as we know, inaccurate and prone to serious flaws when observing people of colour, reflecting the way in which an algorithmic system will always reflect and magnify the biases of the society in which it is created. No AI-based system can match the abilities of human 'super-recognisers' (Bobak et al 2019; Moshakis 2018), but humans tire; the future must lie in a properly-constructed machine learning and human-machine teaming to create an effective and accurate 'always-on' recognition system, which is always overseen by humans. When Cummings praises Petrov while simultaneously calling for a greater use of AI, he forgets that his hero is a perfect example of a human intelligence rejecting the opinion of a machine intelligence. A truly cybernetic approach recognises that what matters is neither human nor machine intelligence per se, but *intelligence* pure and simple. To return to the foundations of the discipline, we must be guided by Wiener (1989):

*Whether we entrust our decisions to machines of metal, or to those machines of flesh and blood are bureaus and vast laboratories and armies and corporations, we shall never receive the right answers to our questions unless we ask the right questions (p. 185-6)*

## References

- Ampère, A-M. (1834). *Essai sur la philosophie des sciences, ou, Exposition analytique d'une classification naturelle de toutes les connaissances humaines*, Bachelier, Paris.
- Arquilla, J. and Ronfeldt, D. eds. (1997a). "Cyberwar is Coming!", in Arquilla and Ronfeldt (1997b), pp.23-60.
- Arquilla, J. and Ronfeldt, D. eds. (1997b). *In Athena's Camp: Preparing for Conflict in the Information Age*, Santa Monica, RAND.
- Beer, S. (1973). "The Real Threat to "All We Hold Most Dear"", *Designing Freedom: notes in support of Lecture 1*, [transcript], accessed at [https://monoskop.org/images/e/e3/Beer\\_Stafford\\_Designing\\_Freedom.pdf](https://monoskop.org/images/e/e3/Beer_Stafford_Designing_Freedom.pdf)
- Bousquet, A. (2008a), "Chaoplex warfare or the future of military organization". *International Affairs*, vol. 84, issue 5, (September), pp. 915-29.
- (2008b). *The Scientific Way of Warfare: Order and Chaos on the Battlefields of Modernity*, Hurst, London.
- Cebrowski, Arthur K. and John J. Garstka. (1998) "Network-Centric Warfare: Its Origin and Future." *U.S. Naval Institute Proceedings*, vol. 124, no. 1 (January), pp.28-35.
- Clarke, Richard A. (2010). *Cyber War*, HarperCollins, London.
- Clynes, M. E., and Kline, N.S. (1960). "Cyborgs and Space", *Astronautics*, September, pp, 26-7 and 74-6.
- Cummings, D., et al (2020a). "The highlights of history: a Spectator Christmas survey". *The Spectator*, 19 December. [online]. [www.spectator.co.uk/article/the-highlights-of-history-a-spectator-christmas-survey](http://www.spectator.co.uk/article/the-highlights-of-history-a-spectator-christmas-survey).
- Cummings, D. (2020b). "'Two hands are a lot' — we're hiring data scientists, project managers, policy experts, assorted weirdos...". [dominiccummings.com](http://dominiccummings.com). [online]. [dominiccummings.com/2020/01/02/two-hands-are-a-lot-were-hiring-data-scientists-project-managers-policy-experts-assorted-weirdos/](http://dominiccummings.com/2020/01/02/two-hands-are-a-lot-were-hiring-data-scientists-project-managers-policy-experts-assorted-weirdos/).
- Ellis, W. (w), Fabry, G. (p and i), and Sharp, L. (i). (2003). *Big Wheel*. Global Frequency. Issue 2, January. Wildstorm, La Jolla.

**Keith Scott**

- Emanuel, P., Walper, S., DiEuliis, D., Klein, N., Petro, J. B. and Giordano, J. (2019). *Cyborg Soldier 2050: Human/Machine Fusion and the Implications for the Future of the DoD*, U.S. Army Combat Capabilities Development Command Chemical Biological Center, Aberdeen Proving Ground, MD.
- Fogleman, R. R. (1995). "Information Operations: The Fifth Dimension of Warfare", *Defense Issues*, vol. 10, no. 47, [online], [www.hsdl.org/?view&did=439942](http://www.hsdl.org/?view&did=439942)
- Galeotti, M. (2016). "Hybrid, ambiguous, and non-linear? How new is Russia's 'new way of war'?", *Small Wars & Insurgencies*, 27:2, pp. 282-301.
- Galeotti, M. (2019). "The mythical 'Gerasimov Doctrine' and the language of threat". *Critical Studies on Security*, 7, pp. 157 - 161.
- Gerasimov, V. (2016). *The Value of Science Is in the Foresight: New Challenges Demand Rethinking the Forms and Methods of Carrying out Combat Operations*. *Military Review* (January-February), pp. 23-9. Available at: [www.armyupress.army.mil/portals/7/military-review/archives/english/militaryreview\\_20160228\\_art008.pdf](http://www.armyupress.army.mil/portals/7/military-review/archives/english/militaryreview_20160228_art008.pdf)
- Green, J.A. (1981). *Cyber warfare : a multidisciplinary analysis*, Routledge, London.
- Krepinevich, A. (2010). *7 Deadly Scenarios: A Military Futurist Explores War in the Twenty-First Century*, Bantam, London.
- Lewis, J. (2018). *The 2020 Commission Report on the North Korean Nuclear Attacks Against The United States*. W.H. Allen, London.
- Lin, P. (2012). "More Than Human? The Ethics of Biologically Enhancing Soldiers". *The Atlantic*, 16 February. [online]. [www.theatlantic.com/technology/archive/2012/02/more-than-human-the-ethics-of-biologically-enhancing-soldiers/253217/](http://www.theatlantic.com/technology/archive/2012/02/more-than-human-the-ethics-of-biologically-enhancing-soldiers/253217/).
- Marks, R. J. (2020). "Meet the U.S. Army's New Drone Swarms". *mindmatters.ai*. [online]. 11 September, <https://mindmatters.ai/2020/09/meet-the-u-s-armys-new-drone-swarms/>.
- McFate S. (2019b). *Mercenaries and War: Understanding Private Armies Today*. NDU Press, Washington.
- McFate, S. (2019a). *Goliath: Why the West Doesn't Win Wars. And We Need to Do About It*, Penguin, London.
- Moore, C. (2019). *Russia And Disinformation: The Case Of Ukraine*, CREST, London.
- Moshakis, A. (2018). "Super recognisers: the people who never forget a face". *The Guardian*, 11 November. [online]. [www.theguardian.com/uk-news/2018/nov/11/super-recognisers-police-the-people-who-never-forget-a-face](http://www.theguardian.com/uk-news/2018/nov/11/super-recognisers-police-the-people-who-never-forget-a-face).
- Newman, J., and Unsworth, M. (1984). *Future War Novels: An Annotated Bibliography*, Oryx Press, Phoenix.
- Palmer, P. (2012). *Red and Blue: Terror*, radio programme, BBC Radio 4, 25 April.
- Qiao, L., and Wang, X. (2020). *Unrestricted Warfare: China's Master Plan To Destroy America*, Shadow Lawn Press, Lambertville [Kindle Edition].
- Ramon, M., Bobak, A. K., & White, D. (2019). Super-recognizers: From the lab to the world and back again. *British journal of psychology*. 110 (3), 461–79.
- Royal Society. (2012). *Brain Waves Module 3: Neuroscience, conflict and security*, Royal Society, London.
- Safi, M. (2019). "Are drone swarms the future of aerial warfare?". *Guardian*, 4 December. [online]. [www.theguardian.com/news/2019/dec/04/are-drone-swarms-the-future-of-aerial-warfare](http://www.theguardian.com/news/2019/dec/04/are-drone-swarms-the-future-of-aerial-warfare).
- Safi, M., Ellis-Petersen, H., and Davidson, H. (2020). "Soldiers fell to their deaths as India and China's troops fought with rocks". *The Guardian*. 17 June. [online]. [www.theguardian.com/world/2020/jun/17/shock-and-anger-in-india-after-worst-attack-on-china-border-in-decades](http://www.theguardian.com/world/2020/jun/17/shock-and-anger-in-india-after-worst-attack-on-china-border-in-decades).
- Shakaria, P., Shakarian, J., and Ruef, A. (2013). *Introduction to cyber-warfare : a multidisciplinary approach*, Elsevier, Amsterdam.
- Singer, P.W., and Cole, A. (2015). *Ghost Fleet: A Novel of the Next World War*, Houghton Mifflin, New York.
- Slaughterbots. (2017). [Online]. Dir. Stewart Sugg. USA: Space Digital/Future of Life Institute. Available on YouTube: [www.youtube.com/watch?v=9CO6M2Hs0IA](http://www.youtube.com/watch?v=9CO6M2Hs0IA).
- Taddeo, M. (2012). "An analysis for a just cyber warfare," *2012 4th International Conference on Cyber Conflict (CYCON 2012)*, Tallinn, pp. 1-10. Accessed at: <https://ieeexplore.ieee.org/document/6243976>.
- Wiener, N. (1948) *Cybernetics: Or Control and Communication in the Animal and the Machine*, Hermann & Cie, Paris.
- Wiener, N. (1989). *The Human Use of Human Beings*, Free Association Books, London. (first published 1950).

# Emergency Response Model as a Part of the Smart Society

Jussi Simola<sup>1,2</sup>, Martti Lehto<sup>1</sup> and Jyri Rajamäki<sup>2</sup>

<sup>1</sup>University of Jyväskylä, Finland

<sup>2</sup>Laurea University of Applied Sciences, Finland

[jussi.hm.simola@jyu.fi](mailto:jussi.hm.simola@jyu.fi)

[martti.j.lehto@jyu.fi](mailto:martti.j.lehto@jyu.fi)

[jyri.rajamaki@laurea.fi](mailto:jyri.rajamaki@laurea.fi)

DOI: 10.34190/EWS.21.079

**Abstract:** Centralized hybrid emergency model with predictive emergency response functions are necessary when the purpose is to protect the critical infrastructure (CI). A shared common operational picture among Public Protection and Disaster Relief (PPDR) authorities means that a real-time communication link from the local level to the state-level exists. If a cyberattack would interrupt electricity transmission, telecommunication networks will discontinue operating. Cyberattack becomes physical in the urban and maritime area if an intrusion has not been detected. Hybrid threats require hybrid responses. The purpose of this qualitative research was to find out technological-related fundamental risks and challenges which are outside the official risk classification. The primary outcomes can be summarized so that there are crucial human-based factors that affect the whole cyber-ecosystem. Cybersecurity maturity, operational preparedness, and decision-making reliability are not separate parts of continuity management. If fundamental risk factors are not recognized, technical early warning solutions become useless. Therefore, decision-makers need reliable information for decision-making that does not expose them to hazards. One of the primary aims of hybrid influence is to change political decision-making. Practically, this means a need to rationalize organizational, administrative, and operative functions in public safety organizations. Trusted information sharing among decision-makers, intelligence authorities, and data protection authorities must be ensured by using Artificial Intelligence (AI) systems. In advanced design, protection of critical infrastructure would be ensured automatically as part of the cyber platform's functionalities where human-made decisions are also analyzed. Confidential information sharing to third parties becomes complicated when the weaknesses of crucial decision-making procedures have been recognized. Citizens' confidence in the intelligent system activities may strengthen because of the decision-making process's reliability. Existing emergency response services are dependent on human ability.

**Keywords:** critical infrastructure protection, cyber ecosystem, emergency response, public protection, and disaster relief, artificial intelligence

---

## 1. Introduction

As earlier researches (Simola & Rajamäki, 2015; Simola & Rajamäki, 2017) has shown, technical solutions need a deeper understanding of user needs. That means the infrastructure of a smart city environment cannot be developed separately from user requirements. There is also a need to design a common emergency response ecosystem for European public safety actors. Therefore, communication solutions used within public safety authorities must suit well in urban and rural areas.

Public safety actors like European law enforcement agencies need a common shared situational picture for the cross-bordering tasks so that operational cooperation is based on a reliable platform. Formal integration in the European Union and between member countries has developed rapidly. That does not mean that collaboration between organizations has developed in the same proportion. Digitalization cannot evolve in isolation from society. There are fundamental needs within public European safety organizations that should be at the same level in every country.

Decision-makers in Finland need to consider that cybersecurity maturity, operational preparedness, and decision-making reliability are integral parts of continuity management. Technical early warning solutions become useless to develop if crucial risk factors are not detected. Therefore, decision-makers need reliable decision-support information for decision-making that does not expose them to hazards. Technological development, infrastructure development, and legislation changes are inner-country challenges and everyday European needs concerning safety development agendas. State-level factors should be added to the European safety framework. There are many strategic plans at the European level concerning safety functions, but national implementation realizes in a different order. As the report of the SAI (2017) indicates, Finland has a lot to do to improve the information exchange in significant accident situations.

Citizens choose political decision-makers, but the highest authorities are selected on selection criteria. Hybrid influencing can destabilize society in many ways, especially if threats accumulate or arise from within the society (Simola, 2020). One of the primary key aims is to influence political decision-making. In practice, this means a need to rationalize organizational, administrative, and operative functions (SAI, 2017). The flow of reliable information between decision-makers, intelligence authorities, and data protection authorities must also be ensured by using artificial intelligence systems. In an ideal model, national protection of vital functions would be ensured automatically as part of the cyber platform's functionalities where human-based decisions are also analyzed. When human weaknesses are left out of decision-making procedures, e.g., data leakage to third parties becomes more difficult. It could increase citizens' confidence in the smart system's activities and increase trust in government institutions.

Security and intelligence agencies in Europe have acquired new rights under the law. In Finland acceptance of Intelligence legislation package concerning civilian and military intelligence legislation has been approved. It will be seen in the future how prepared our state-level decision-makers are to develop the legislative base for the new cyber-physical ecosystem. A substantial part of Finland's intelligence legislation has been updated to the same level as in other European countries. The rest of this paper is divided as follows. Section 2 handles the overview of the theoretical framework. Section 3 proposes the central concepts of critical infrastructure and the framework of this article. Section 4 presents the research background, objectives, and methods. Section 5 presents the findings. Section 6 includes a discussion about the research area. Section 7 handles conclusions.

## **2. Theoretical framework and literature review**

Member countries of the European Union and smart cities need cooperation because, without smart cities, the European Union's intelligent ecosystem cannot be created. Financial competition between countries creates the need for the development of intelligent technology. Thus, intelligent information systems are being developed; there must be an already digital ecosystem to connect the system. Every smart city should be constructed from a long-term view. A smart city needs an urban built technology-oriented environment where different kinds of intelligent systems communicate with each other. This case study aims to find out those fundamental technological-related risks that expose society to hybrid threats. These threats affect the protection of critical infrastructure and prevent the detection of threats. Implementation of the presented Hybrid Emergency Response Model is the primary purpose because there are separate situation centers, emergency response centers, and organizations fighting against cyber threats. Still, there is no common emergency response model for all kinds of hybrid-threats. The main author of this research has innovated the next-generation emergency response model (Simola & Rajamäki, 2017).

### **2.1 Development of Emergency Response system solutions**

Emergency Response Center uses an Emergency Response system. It is one kind of decision support system. Decision support systems are used to track key incidents and the progress of responding units, optimize response activities and act as a mechanism for queuing ongoing incidents (Ashish et al., 2007; Endsley, 1988; Endsley, 1995).

In Finland, traditional emergency response functions have been modeled from other countries. However, we still have significant challenges related to the possibilities of transferring emergency data correctly and in time to the Emergency response center. There was a separate emergency response unit in the Police organizations until 1999. E.g., regional Radio Police consisted of their dispatch personnel who answered citizens' emergency calls and managed the use of emergency units to the site of an accident. Also, municipal rescue services handled their emergency calls. In the 21<sup>st</sup> century, separate emergency call units and functions were combined with emergency response centers. Very soon after the organization's changes, PPDR authorities found the need to manage their emergency resources. PPDR organizations established their situation centers to allocate emergency resources concerning field workers' cooperation.

The culture of the organization needs to be understandable when the purpose is to develop new technological solutions. Public safety organizations have a common working culture but also separate inner-organizational subcultures. That same issue concerning the meaning of the working culture relation to organizational reform also occurs in a different atmosphere and a different field. In practice, smart city infrastructure is the fundamental framework that governs minor factors inside it. It is impossible to create technological solutions in their separate entity regardless of the organizations' culture.

## **2.2 Smart nations and smart cities**

Political power relations affect the national future of digitalization. Urbanization changes our lifestyle, and the digitized environment creates the base for the new safety culture. Citizens meet friends in public places, and they might go to the shopping center for shopping goods. Time has changed more dangerous; global terrorism has impacted people's behavior. Historical similarities between countries in northern Europe helps to understand the safety needs of neighboring countries. While separate European societies are evolving, societies are developing their cooperation on digitalization. It is essential to see the digitalization development of the north from the same perspective. There are different political aspects between European Union countries concerning energy and security policy. EU as the commercial operator brings its own needs into the discussion. Collaboration with Russian and China challenges our culture and western way of thinking. We need cooperation, but possibilities for cybersecurity threats emerge too often (Robertson & Riley 2018). Nord Stream2 and different kinds of 5G and cable projects may expose national security under cross-bordering hybrid risks (Buchanan, 2017; Shackelford et al., 2017; Buchanan, 2018; Hutchens, 2018).

It is impossible to create the entirety of a smart society without understanding the continuity management of society. If departments of the central government design separate digitalization projects without a common understanding of the future needs, society's expenses and digitalization management become complex. The governance of digitalization needs common goals for all participants. It means that the regional and local administrative operators need exact central steering concerning all municipal constructions of infrastructure.

## **3. Critical infrastructure**

The United States define critical infrastructure as physical or virtual systems and assets that are so vital that destructions of the above would have a crucial influence on security, national economic security, national public health, and safety, or any combination of those matters (The White House, 2013). According to the Secretariat of the Security Committee (2013), critical infrastructure comprises vital physical facilities, infrastructures, and electronic functions and services.

Critical Information Infrastructure comprises any physical or virtual information system that controls, processes, transfers, receives, or stores electronic information in any form, including data, voice, or video that is vital to the functioning of critical infrastructure (DHS 2011).

### **3.1 Fundamental elements of critical infrastructure in smart society**

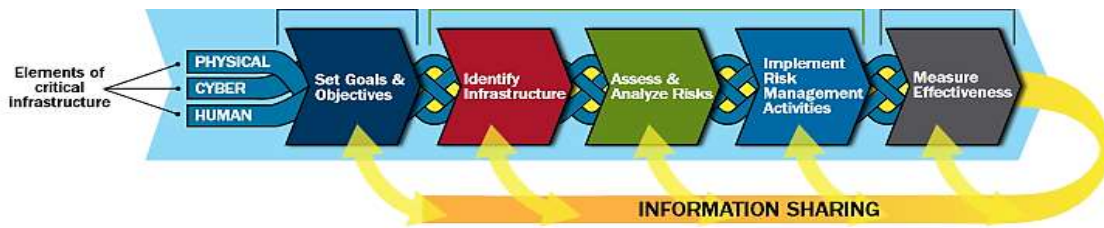
U.S. Department of Homeland Security (2013) classifies 16 different sectors for the Critical Infrastructure as follows: "Chemical, Commercial Facilities, Communications, Critical Manufacturing, Dams, Defense Industrial Base, Emergency Services, Energy, Financial Services, Food and Agriculture, Government Facilities, Healthcare, and Public Health, Information Technology, Nuclear Reactors, Materials and Waste, Transportation Systems and Water Wastewater System" (DHS 2013).

Department of Homeland Security categorizes, e.g., the communication sector closely linked to the Energy sector, the Information sector, the Financial services, the Emergency services, and the Transportation system sectors (DHS, 2013). Every government uses a different emphasis level between the importance of emphases. In this research communication sector, the energy sector, information technology, and emergency services sector have been chosen as selected sectors of critical infrastructure.

### **3.2 Risk management and preparedness**

According to (NIST, 2018) the framework is used in U.S. suites well also in Finland. The risk management framework consists of three elements of critical infrastructure (physical, cyber, and human) that are explicitly identified and should be integrated throughout the steps of the framework. The critical infrastructure risk management framework supports a decision-making process that critical infrastructure actors or partners collaboratively undertake to inform the selection of risk management actions. It has been designed to provide flexibility for use in all sectors, across geographic regions, and by various partners. It can be tailored to dissimilar operating environments and applies to all threats (DHS, 2013).





**Figure 1:** Critical infrastructure risk management framework

The risk management concept enables the critical infrastructure actors to focus on those threats and hazards that are likely to cause harm and employ approaches that are designed to prevent or mitigate the effects of those incidents. It also increases security and strengthens resilience by identifying and prioritizing actions to secure continuity of essential functions and services and support enhanced response and restoration (DHS, 2013).

According to the Department of Homeland Security (2013), the first point recommends setting infrastructure goals and objectives that are supported by objectives and priorities developed at the sector level. To manage critical infrastructure risk effectively, actors and stakeholders must identify the assets, systems, and networks that are essential to their continued operation, considering associated dependencies and interdependencies. This dimension of the risk management process should also identify information and communications technologies that facilitate essential services (DHS, 2013).

The third point recommends assessing and analyzing risks. Those Risks may comprise threats, vulnerabilities, and Consequences. A threat can be a natural or human-made occurrence, individual, entity, or action that has or indicates the potential to harm life, information, operations, the environment, and/or property. The vulnerability-based risk may occur physical feature or operational attribute that renders an entity open to exploitation or susceptible to a given hazard. A consequence can be the effect of an event, incident, or occurrence. Implementing risk management activities means that decision-makers prioritize activities to manage critical infrastructure risk based on the criticality of the affected infrastructure, the costs of such activities, and the potential for risk reduction. The last element measuring effectiveness means that the critical infrastructure actors evaluate the effectiveness of risk management efforts within sectors and at national, state, local, and regional levels by developing metrics for both direct and indirect indicator measurement (DHS, 2013).

In this research, we have used a modified combination of NIST and Octave Allegro Risk Assessment Frameworks. According to Caralli & al. (2007), Octave allegro is a strategy for prioritizing and sharing information about security risks, e.g., information technology. According to (Zio & Pedroni, 2012) NASA risk-informed risk is the potential for performance shortfalls, which may be realized in the future, with respect to achieving explicitly established and stated performance requirements. As Figure 2 illustrates, Risk Management by NASA integrates two complementary processes, Risk-Informed Decision Making (RIDM) and Continuous Risk Management (CRM), into a single coherent framework. The RIDM process addresses the risk-informed selection of decision alternatives to assure effective approaches to achieving objectives, and the CRM process addresses the implementation of the selected alternative to ensure that requirements are met. These two processes work together to assure effective risk management as NASA programs (NASA, 2015).



**Figure 2:** Combined risk management processes

### 3.3 Protecting vital society

According to (The Security Committee, 2018; Ministry of defence, 2010), threats can occur on the individual, national, and global levels. Individual threats primarily affect the individual, national threats primarily affect the state, society and population, global threats affect the earth and the population's future security. Figure 3 illustrates those levels relations. According to the Ministry of the Interior (2018) three top-level threat scenarios

are severe disturbances in the power supply and cyber threats like severe disturbances in the telecommunications and information systems. Vital functions to the Finnish society contain the management of Government affairs, international and EU activities, Finland's defence capability, internal security, the functioning of the economy, infrastructure and security of supply, functional capacity of the population and services and psychological resilience to a crisis (Ministry of the Interior, 2018).



Figure 3: Threats on the individual, national, and global level

### 3.4 Artificial Intelligence helps continuing management

Artificial Intelligence (AI) is a part of the system that displays intelligent behavior by analyzing their environment and taking multiple actions with autonomy to achieve given purposes. Software-based artificial intelligence systems can act in the virtual world consisting of image analysis software and search engines. Also, it may be embedded in hardware devices, e.g., advanced robots, unmanned vehicles, or Internet of Things applications (European Commission 2018).

An intelligent Agent (IA) is an entity that produces decisions. It allows performing, e.g., specific tasks for users or applications. It can learn during the process of performing tasks. Two main functions consist of perception and action. Intelligent Agents form a hierarchical structure that comprises different levels of agents. A so-called multi-agent system consists of several agents that interact with one another (Wooldridge 2009). That combination may solve challenging problems in society. The agent may behave in three ways: reactively, proactively, and socially (Wooldridge 2009).

## 4. Research background, objectives, and methods

There have been many state-level discussions concerning digitalization among decision-makers in media. At present public safety authorities and decision-makers do not use cyber-threat information in their operative daily routine almost at all. The challenge is that public safety authorities have separate cybersecurity organizations in their administrations. Organizations that have responsibilities for cybersecurity operations act as separated entities from PPDR services. As a part of TRAFICOM, the National Cyber Security Centre Finland (NSCS-FI) produces and shares cyberthreats information for stakeholders. Still, shared data does not achieve emergency response centers or situation centers. Separate organizational cybersecurity functions, methods, and procedures prevent an effective response to cyber-physical threats. In addition to this, developed innovations, e.g., emergency response systems, are all useless if our ministers and other decision-makers are not faithful or decisions are made to advantage a foreign power. It is essential to realize the source and degree of threat. The innovative urban areas and information systems may be constructed on an unstable ground level that may consist, e.g., energy supply solutions and dicey communication equipment. Overall situational awareness enhances by combining Open Source Intelligence data and traditional intelligence data (Morrow and Odierno 2012). The cyber situational picture is needed because Hybrid threats need hybrid responses.

### 4.1 Method and process

The multimethodological approach consists of four case study research strategies: theory building, experimentation, observation, and systems development (Nunamaker & al., 1990). Yin (2014) identifies five components of research design for case studies: (1) the questions of the study; (2) its propositions if any; (3) its unit(s) of analysis; (4) the logic linking the data to the propositions; and (5) the criteria for interpreting the findings. This research is carried out with the guidance of Yin (2014). The research concentrates on sources of

scientific publications, collected articles and literary material. The research subject comprises public safety organizations, procedures, and vital functions of Finland society.

The first purpose of this qualitative research was to collect and classify selected risks from different risk areas. In this research, we have used the Modified Risk Assessment Framework. The second purpose was to find out hidden technological-related state-level risks and challenges that are outside the official risk classification. A simple process model helps to identify those fundamental factors that are used in the creation of the scenarios. We have defined the research area concerning vital functions in four main sections; the Emergency services sector, the Communication sector closely linked to the Energy Sector, and the Information sector. Firstly, it is essential to find out technological-related risks and scenarios that expose society's vital functions to hybrid-threats and risks. It is easier to detect fundamental level risk factors when basic threats and risks are categorized and classified. These threats affect the protection of vital functions and prevent the detection of threats. We have used a combination of different methodologies to find out those factors that affect decision-making in society. As Table 1 illustrates, separate risks are divided into the main areas as follows: Administrative risks, conflict risks, emergency functions related risks, socioeconomic risks and infrastructure-related risks. The numbers A, B, C, D, and E indicate which main category the subcategories are also linked. Separate risks are categorized and ranked on a three risks level process. The first measure is valued "frequency of the phenomenon" (1 = phenomenon does not occur every year, 2 = phenomenon occurs yearly, and 3 = a phenomenon is permanent). The second value is titled "predictability and measurability of risks" (1= phenomenon is neither predictable nor measurable, 2= phenomenon is predictable. 3 = phenomenon is predictable and measurable.) The third value is named "impact of risk on overall security" (1= impact of the risk on one vital function, 2=impact of the risk on two to three vital functions, and 3 = impacts of risk to more than three selected vital functions.) Coefficients for variables are 1 to "frequency of the phenomenon," 2 to predictability and measurability of risks, and 3 to "Impact of risk on overall security.

**Table1:** Main risk classification

Main risk classification and subcategories							
A		B		C		D	E
Administrative risks		Conflict risks		PPDR services and functions related risks		Socioeconomic risks	Infrastructure related risks
Problems in local continuity management	C,D	Cyberattacks	A,C,E	Overloaded Emergency management system	B,E	Unemployment	Structural problems in the built urban area
Problems in cooperation between decisionmakers	B,C,D,E	Human made disasters or pandemics	F	Lack of human resources in PPDR services	A,D,E	Refugees	Structural problems in the rural area
Separate municipal activities	E	Cross-border radiation	C,D,E	Lack of resources in PPDR services/	A,D,E	Cultural change	Recovery problems
Organizational problems	B,C	Physical war	A,C,D,E	Emergency event	D,E	Use of substances	Secrets cyber influences
Leadership problems in government	B,C,D,E	Hybrid warfare	A,C,D,E	Resource awareness of volunteers	A,D,E	Citizens poverty	Communication problems
						Unidentified people	

The research aims to create a decision support subsystem solution for the proposed Hybrid Emergency Response system to assist politicians and public sector actors. That is an important issue because there is a need to detect sources of threats much earlier.

We have used the methodology model and framework by the National Aeronautics and Space Administration in designing the subsystem of Hybrid emergency response systems. The continuous Risk Management (CRM) process stresses the management of risk during implementation. The Risk-Informed Decision Making (RIDM) methodology is part of a systems engineering process that emphasizes the proper use of risk analysis in its broadest sense to make risk-informed decisions that impact all mission execution domains, including safety, technical, cost, and schedule. RIDM helps ensure that decisions between alternatives are made with an awareness of the risks associated with each helping to prevent late design changes, which can be key drivers of risk and cancellation (NASA, 2016).

Figure 4 illustrated the risk analysis framework that helps to analyze the different alternatives and factors when decision-makers are making final decisions (Dezfuli et al.,2010; Zio and Pedroni, 2012).

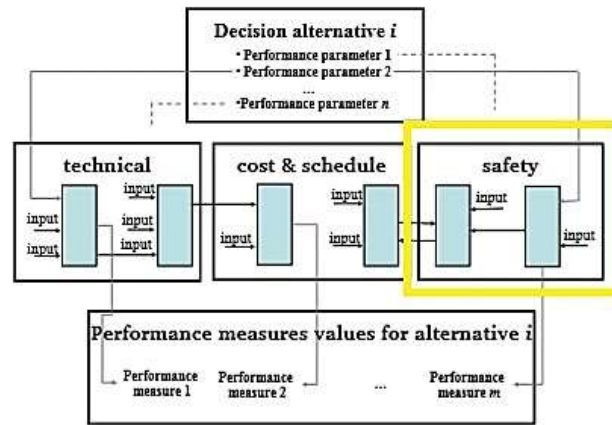


Figure 4: Risk analysis framework

The study's main goal is to find out fundamental societal factors that affect the effective protection of critical infrastructure. This research divides the types of risks into four sections. Ground Level indicates fundamental risks with scenarios that include factors, events, and actions of society. The scenarios' essential factors put all other societal factors, events, or actions into secondary threats level. Fundamental factors also make it possible to realize lower-level threats. This causes that the effective protection of critical infrastructure depends on external factors. The operator who controls external factors also dominates critical infrastructure. Therefore fundamental ground-level risk factors should be recognized and minimized.

## 5. Findings

Table 2 illustrates elements of society between risk levels. Higher risk levels are on the right, and these elements set the greatest threats to the vital functions. If ground-level threats are realized, the protection of critical infrastructure loses its meaning. E.g., the wide use of substances may indirectly harm society's overall security, but addiction cannot remain hidden for a long time. As a member of the EU, Finland gave away part of the national parliaments' sovereignty concerning national regulation. This kind of problem may happen when supranational legislation gives away the power of decision-making from the government to the commercial operators. E.g., change of ownership of the electricity transmission network.

Table 2: Classifications and impacts of risks

Classified by effectiveness of fundamental hidden risks and scenarios (red level) - Impacts and disruption on selected scenarios and consequences. Level of risks based on three values (frequency of the phenomenon, predictability and measurability of risks and impact of risk on overall security). Impact level 1-3 (1=low, 2=average level, 3=high impact) 1 = impact on 1-2 scenarios, 2 = impact on 3-4 scenarios, 3 = impact on 5-6 selected scenarios. 1 = X, 2 = XX, 3 = XXX							
Classified basic risk levels. 1=low 4=high	1	2		3	4		
	levels 6-10 =1	levels 11-13 =2		levels 14-16 =3	levels 17-18 = 4		
Refugees	X	Overloaded Emergency management system	XX	Structural problems in the rural area	XX	Cyberattacks	XXX
Cultural change	X	Lack of resources in PPDR services/	XX	Human made disasters or pandemia	XX	Separate municipal activities	XXX
Use of substances	X	Resource awareness of volunteers	X	Structural problems in the built	XXX	Secrets cyber influences	XXX
Unemployment	X	Emergency event	X	Leadership problems in government	XXX	Hybrid warfare	XXX
		Cross-border radiation	X	Lack of human resources in	XX	Unidentified people	XXX
		Organization al problems	XX	Communication problems	XXX		
		Problems in local continuity management	XX	Problems in cooperation between decision makers	XXX		
		Citizens poverty	X	Recovery problems	XXX		
				Physical war	XXX		

Findings indicate that lower-level risks of critical infrastructure do not cause problems to the ground-level risks. Higher-level risks also indicate structural governance problems in society. The effectiveness level indicates threats' impacts to the vital functions. Three x means that basic independent level risk becomes more dangerous due to connection ground level scenarios. As Table 3 illustrates, six scenarios were selected. At which impact level selected risks to affect to potential consequences of the scenarios? As illustrated in table 1 one X indicate impact on 1-2 scenarios, XX indicate impact on 3-4 scenarios, XXX = indicate impact on 5-6 selected scenarios. If higher (4) level risk support 4 or more scenarios and consequences, impact level is occasional for all vital functions. The domino effect causes this change of situation. E.g., a separate cyberattack is not so dangerous, but the event's danger will essentially change if it is due to a political decision.

**Table 3:** Scenarios and consequences

<b>Ground-level - Scenario</b>	<b>Consequences</b>
A) Legislation – Lack of possibilities to intervene in internal security	Lack of internal self-determination and internal sovereignty
B) Political decisions – Lack of continuity	Line changes in security policy – development of unstable decision-making culture
C) Energy solutions – Dependence on imported energy management, short-term political purposes	Exposure to extortion by an external actor
D) Equipment for Communication systems – E.g., 5G solutions devices, network equipment	Foreign state spying and foreign country get a role in infrastructure
E) International public projects – Smart cable projects, gas pipeline projects	Vulnerability to sabotage – the foreign state may use cables and pipelines for hybrid influencing

Threats like severe disruptions to a power supply, severe disruptions to telecommunications and information systems risks are noticed in Finland's security strategy for society report. Still, the same fundamental risk types occur as the causes that have not been considered in decision-making.

## 6. Discussion

In Finland, existing solutions for public operators based on outdated technology and systems' life-cycles are short (DHS, 2018). Currently, the victim of an accident may have to wait long for the emergency response center's response because call center personnel have to exercise how the new Emergency Response Center system works (Saarenpää J. & Virtanen V. 2019). The handling of incoming and outgoing phone calls will lengthen.

Development towards the digital ecosystem starts with cultural understanding and process management. The subcultures of different PPDR authorities should be implemented through systems. Currently, all actors have their own separate operating model. E.g., if a complete emergency response system requires a significant additional workforce, designing has failed. Technological opportunities have not been exploited in Finland, such as in the U.S. The introduction of an immature system on holiday does not reflect the understanding of the situation in the operating environment (Rahko, 2018). A fully automated emergency response center can be a reality within a decade. An automated decision support system for the highest decision-makers can be a reality soon because vital functions require proof of political decisions.

## 7. Conclusions

As discussed above, we cannot hide our history and culture, but if we are developing a cyber-secure smart ecosystem, we need to make changes to the decision-making culture. The research has been shown that different kinds of structural fundamental-level threats may occur before any classified threat has been illustrated. Engineers, architects, and designers cannot develop anything new concerning smart solutions if the ground base is weak. An unsecured platform causes fundamental obstacles to designing solutions for an intelligent society. Legislation set challenges to the national politicians and authorities, but also power relations between union countries.

The micro and macro levels will be encountered if a foreign state party intervenes to interfere with data traffic functioning in maritime areas. E.g., there is a northeast cable project designed to connect networking activities between different continents. Nowadays, the problem is that fiber optic and power supply are transmitted through the same cable. So-called unexpected happenings influence all ecosystems. This kind of threat comes true and happens out of public safety control. In the future, it is an occasional issue to find the right balance between national security and warm bilateral relations.

Vulnerabilities and risks have increased, though formally, the goal is to harmonize Eastern and Western data cable functionalities (Buchanan, 2018; Shackelford et al., 2017). The study shows that the most troublesome and most significant threats to national security and vital functions are related to human factors, that are based on politicians' decisions and political projects. It is challenging to anticipate national policy's real direction at the macro level because good inter-state relations may indicate ignoring security issues. The study suggests that artificial intelligence-based solutions should be used enhancing to support decision-making. The subsystem could also operate as a part of the next-generation emergency response model. This model will work in two ways. Firstly, the framework consists of predictive and preventive elements that react when cyber-threat data fusion produces signals through the AI-agents and sensors that activate actuators, e.g., bollards or evacuation systems in smart cities infrastructure. Secondly, the system will output handled data for the decision-makers as politicians. This dimension uses the method that connects small pieces of data into a big view producing the situational picture. At present, state-level political decision-making culture may prevent the proposed smart hybrid emergency model's utilization and usefulness. Decision-makers of Finland need to consider if fundamental risk factors are not recognized, technical early warning solutions become useless.

## References

- Ashish, N., Kalashnikov, D. V., Mehrotra, S., Venkatasubramanian, N., Eguchi, R., Hegde, R., & Smyth, P. (2007). Situational awareness technologies for disaster response. In H. Chen, E. Reid, J. Sinai, A. Silke & B. Ganoz (Eds.), *Terrorism informatics: Knowledge management and data mining for homeland security*. Springer.
- Buchanan, E. (2017). "From Russia with Love: Understanding the Russian Cyber Threat to U.S. Critical Infrastructure and what to do about It." 96 (2).
- Buchanan, E. (2018) Sea Cables in the Thawing Arctic. Lowy Institute, last modified 01.02.2018, accessed 20.08.2018, <https://www.lowyinstitute.org/the-interpreter/sea-cables-thawing-arctic>.
- Caralli R. A., Stevens, J. F., Young, L. R., Wilson, W. R. (2007). *Introducing OCTAVE Allegro: Improving the Information Security Risk Assessment Process*. Technical report. U.S. Software Engineer Institute. Carnegie Mellon University
- DHS, (2011). *Blueprint for a Secure Cyber Future – The Cybersecurity Strategy for the Homeland Security Enterprise*
- DHS, (2013). *NIPP 2013 - Partnering for Critical Infrastructure Security and Resilience*.
- DHS, (2018). *Office of Emergency Communications: Cyber Risks to Next Generations 9-1-1*.
- Dezfuli, H., Stamatelatos M., Maggio G., Everett C., & Youngblood R. (2010). *NASA Risk-Informed Decision Making Handbook: Office of Safety and Mission Assurance NASA Headquarters*.
- Endsley, M. R. (1988). "Design and Evaluation for Situation Awareness Enhancement." *Human Factors Society*.
- Endsley, M.R. (1995). "Toward a Theory of Situation Awareness. *Human Factors*." (37): 32-64.
- European Commission (2018) *Artificial Intelligence for Europe 237*.
- Hutchens, G. 2018 "Huawei Poses Security Threat to Australia's Infrastructure." *The Guardian*, last modified 30.10.2018, accessed 28.02.2019, <https://www.theguardian.com/australia-news/2018/oct/30/huawei-poses-security-threat-to-australias-infrastructure-spy-chief-says>.
- Ministry of Defence. (2010). *Security strategy for society, government resolution*. Helsinki: Ministry of Defence.
- Ministry of the Interior. (2018). *National Risk Assessment 2018*. Helsinki: Ministry of the Interior.
- Morrow, J., & Odierno, R. (2012). *Open-source Intelligence, ATP 2-22.9, army techniques publication*. Washington: Headquarters, Department of the U.S. Army.
- NASA. (2015). *Considering Risk and Resilience in Decision-Making*. Hampton, Virginia: National Aeronautics and Space Administration.
- NASA. (2016). *Systems engineering handbook*. Washington. National Aeronautics and Space Administration.
- NIST. (2018). *Framework for Improving Critical Infrastructure Cybersecurity: National Institute of Standards and Technology*.
- Nunamaker Jr. J., Chen M. & Purdin, T. (1990). *Systems development in information system research*. Vol 7 (3), 89–106.
- Rahko, P. (2018) Uusi tietojärjestelmä otettiin käyttöön Oulun hätäkeskuslaitoksessa onnistuneesti, paikalla oli yöllä lähes kaksinkertainen henkilömäärä. *Kaleva*.
- Robertson, J. and Riley, M. (2018) "The Big Hack: How China used a Tiny Chip to Infiltrate U.S. Companies?" *Bloomberg*, last modified 4.10.2018, accessed 2/28, 2019, <https://www.bloomberg.com/news/features/2018-10-04/the-big-hack-how-china-used-a-tiny-chip-to-infiltrate-america-s-top-companies>.
- Saarenpää J. & Virtanen V. (2019) Erica-hätäkeskustietojärjestelmä Käyttöönoton vaikutukset poliisin päivittäiseen kenttätoimintaan.
- SIA. (2017). *Turku stabbings on 18 August 2017/ Puukotukset Turussa, Safety Investigation Authority, Helsinki 18.8.2017*
- Secretariat of the Security Committee. (2013). *Finland's Cyber Security Strategy - Government Resolution: Ministry of Defense*.
- Shackelford, S. J., Sulmeyer M., Graig Deckard, A. N., Buchanan, B. & Micic, B. (2017). *From Russia with Love: Understanding the Russian Cyber Threat to U.S. Critical Infrastructure and what to do about It*. 96 (2): 321-337.
- Simola J. & Rajamäki J. (2015) "How a real-time video solution can affect to the level of preparedness in situation centers," 2015 Second International Conference on Computer Science, Computer Engineering, and Social Media (CSCESM), Lodz, 2015, pp. 31-36, doi: 10.1109/CSCESM.2015.7331824

***Jussi Simola, Martti Lehto and Jyri Rajamäki***

- Simola, J. & Rajamäki, J. (2017). "Hybrid Emergency Response Model: Improving Cyber Situational Awareness." University, College, Dublin, Ireland, APCI, 29-30 June.
- Simola, J. (2020). Privacy issues and critical infrastructure protection. In: V. Benson and J. McAlhaney, eds, Emerging Cyber Threats and Cognitive Vulnerabilities. Academic Press, pp. 197-226.
- The Security Committee. (2018). Security Strategy for Society. Helsinki: The Security Committee.
- The White House. (2013). Federal register – Improving Critical Infrastructure Cybersecurity
- Wooldridge, M. (2009) An Introduction to Multiagent System, 2 ed. John Wiley & Sons, United States.
- Yin, R. K. (2014). Case Study Research, Design and Methods. 5th ed. Thousand Oaks: Sage Publications.
- Zio, E. and Pedroni, N. (2012). Risk-Informed Decision-Making Process. Toulouse, France: Foundation for an Industrial Safety Culture.



# Joint All-Domain Command and Control and Information Warfare: A Conceptual Model of Warfighting

Joshua Sipper

Air Force Cyber College, Air University, Montgomery, USA

[jasipper@gmail.com](mailto:jasipper@gmail.com)

DOI: 10.34190/EWS.21.018

**Abstract:** A riot of change strategically and operationally has erupted within the joint force, drawing in two powerful concepts: joint all-domain operations (JADO) and information warfare (IW). With renewed emphasis on IW with the cyber-enabled construct including the consequential information related capabilities (IRC) of information operations (IO), intelligence, surveillance, and reconnaissance (ISR), and electromagnetic warfare (EW), and a cross-cutting requirement for joint all-domain command and control (JADC2), the joint force is on the cusp of a significant strategic shift. The following paper and its discussion will explore linkages between IW and JADC2, explain how IW benefits and enables JADC2, and present a conceptual model detailing how IW and JADC2 can work together to produce operational effects and advance US strategic interests now and into the future.

**Keywords:** cyber, information, warfare, intelligence, joint

---

## 1. Introduction

The explosion of capabilities available through technology and the IRCS it drives has become a virtual juggernaut within the auspices of modern warfare theory and practice. IW as a strategic construct has emerged in the twenty-first century as an important way forward in the ongoing struggle of great power competition, especially between the US and China. The IW paradigm is characterized by its four IRCS: cyber operations (CO), ISR, IO, and EW, providing a power scaffold on which to hang and matrix multiple methods of interdisciplinary control methodologies for producing effects across domains and battlespaces. This ability to reach into and across capabilities and domains makes IW a strong foundational strategy for enacting, supporting, and driving the JADC2 operational structure.

The following analysis will begin with a structural examination and aggregation of JADC2 to get a clearer view of the strategic and operational concepts undergirding the construct. Following the JADC2 analysis, the elements of IW will be explored related to how IW affects and is affected by JADC2. The combinatory power of elements will be discussed as well as how each element provides C2 capability and function across the joint domains. Finally, a conceptual model will be introduced, detailing IW IRCS and JADC2 interoperability.

## 2. JADC2 structure and character

Over the centuries, the military profusion and might of the US has grown from a land-centric militia to an army of massive proportions, a privateer sea force to one of global, naval omnipresence, and from balloons used for simple intelligence collection to an air force that controls skies from the ionosphere to the ground. Now, out of the air force, a space force has emerged, furthering the US reach and capabilities beyond the horizon; indeed outside the Earth itself. Finally, with the elevation of cyberspace to domain status, all of the aforementioned domains have only increased in power, influence, and reach. This short history is a picture of the individual elements seen across the world today by enemies and allies alike, but it is not the end of the story. Now, the US military has embarked upon an historical fusion of these domains and the services that oversee and operate within them. “[W]e must integrate our advantages across these domains in new and dramatically effective ways. Linking operations moving at the speed of light with operations moving at the speed of sound requires we bring it all together” (Goldfein, 2018). Through the interleaving of these domains and the force power associated with them, capabilities can be leveraged to produce battlespace effects exponentially greater than they would be individually.

A great deal of the effects sought through JADC2 have to do with the fact that great power competitors like Russia and China have powerfully employed anti-access/area denial (A2/AD) strategies with significant effects. While Russia projects A2/AD into eastern European countries, historically known for Russian ties, China has extended its reach into Taiwan, North Korea, and numerous other Pacific theater regions. “The four Armed Services ... agree that they must conduct operations in all domains, land, sea, air, space and cyber. They even are in general agreement on the initial objective for the Joint Force in such a conflict. This is to penetrate and



disintegrate an adversary's layered and networked arrays of anti-access and area-denial (A2/AD) systems by conducting rapid and coordinated attacks across all domains" (Goure, 2019). Through the aggregation of capabilities amounted in the service elements and domains, effects to counter strategies like A2/AD can be vastly improved.

JADC2 efforts are progressing rapidly across the services with joint exercises underway. Major General Paul Chamberlain confirmed that the Army would be participating in the Air Force's second JADC2 exercise to be held in April but was still determining what their role would be. "[We are waiting for] the development of the Joint warfighting concept, and then we will figure out how we will plug into that" (Underwood, 2020). With training and joint doctrine development moving forward rapidly, the opportunity for bringing capabilities together across all domains is quickly becoming a reality. In fact, the army has already designed the concept (see Figure 1, TRADOC, 2018) with some interesting outcomes. Through the use of integrated capabilities across the domains, power projection of the joint force can drive farther into adversary physical and virtual battlespaces with 1. Competition, 2. Penetration, 3. Dis-integration, 4. Exploitation, and 5. Re-competition.

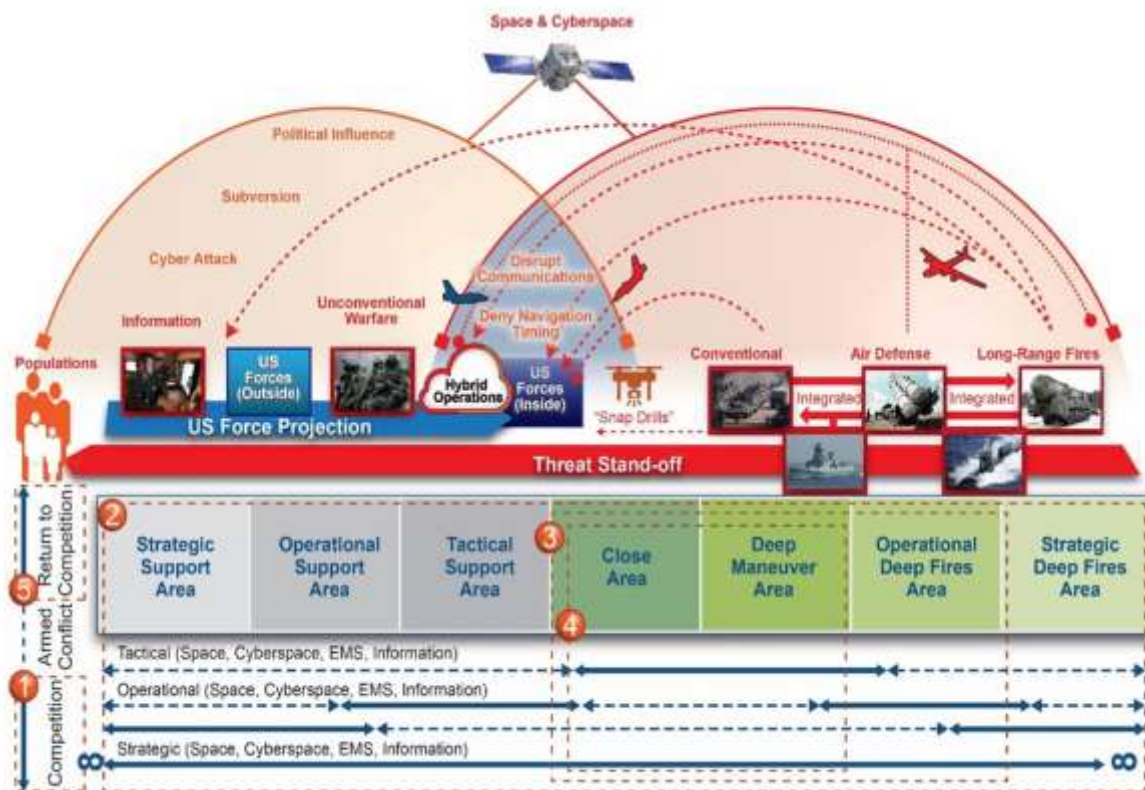


Figure 1: JADO concept

Another vitally important consideration toward the development of a JADC2 strategy and capability is a system designed for the seamless integration of capabilities to enable and create battlefield effects. "The Air Force recently requested \$US435 million for the Advanced Battle Management System (ABMS), the leading technical solution to the problem of Joint All Domain Command and Control (JADC2)" (Dwyer, 2020). With systems being designed specifically for JADC2 implementation, the opportunity for bringing the domains together is even more of a reality. Of course, the integration of these systems is only possible through the use of IRCs and technical capabilities which include the IW construct to be discussed.

The recently signed Air Force JADO Doctrine Note 1-20 offers further advancement of the JADO/JADC2 concepts. As mentioned above, the Army is working closely with the Air Force concerning JADO and the joint concepts that will grow out of this ground-breaking strategic formula. "JADO requires an approach that evolves continuously to take advantage of opportunities as they arise and present a flexible, responsive defense" (Goldfein, 2020). Together, with the capabilities offered through IW, this flexibility and responsiveness can be leveraged to control all domains and exert joint force power at a level many orders of magnitude greater than each strategic concept alone.

General John Hyten gives a solid and penetrating look at what JADO is and what it can do for the US joint force. “All-Domain Operations...combines ‘space, cyber, deterrent, transportation, electromagnetic spectrum operations, missile defense — all of these global capabilities together ... to compete with a global competitor and at all levels of conflict’” (Clark, 2020). This kind of flexibility and capability, coupled with the capabilities and flexibilities of IW offer an even greater push into the physical and virtual, information spaces of adversaries, making it possible for the joint force to map out and control information and actions in increasingly contested and congested environments. “[I]f we figure that out, we’ll have a significant advantage over everybody in the world for a long time” (Clark, 2020). JADO strategy is decidedly not a passing idea, but a persistent and necessary addition to the joint force arsenal of strategies and capabilities; one that can be further enabled through the use of IW and the combined effects (see Figure 2, Orye & Maennel, 2019) and strategies offered there.

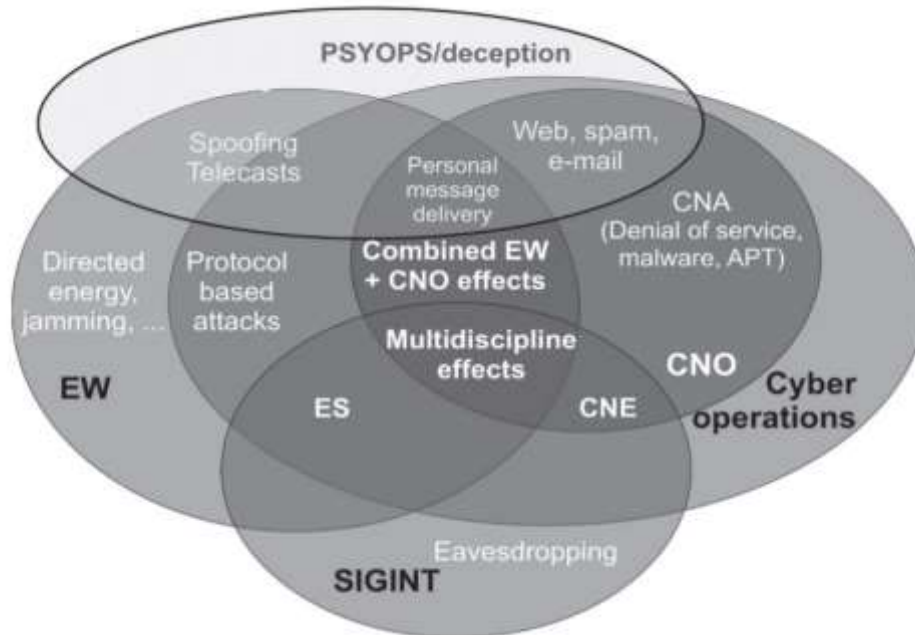


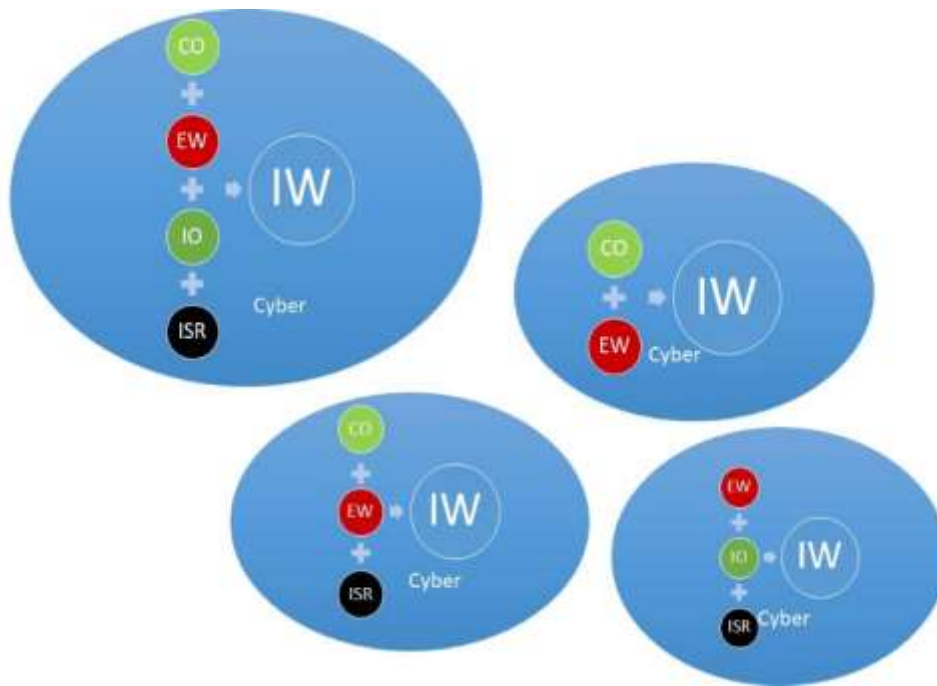
Figure 2: IW and combined effects

### 3. IW IRC elements and interoperability

Although the IRCs of ISR, EW, and IO have been available and in use for most of the 20<sup>th</sup> century and forward, cyber is the latest on the scene and yet the capability that ties the rest of the IRCs together. This unusual placement of an IRC in the company and annals of traditional IRCs makes cyber not only an intriguing field, but a true icon in its classification. When it comes to capability maturity, cyber is definitely a candidate, yet also an ever-growing IRC. This fact simultaneously makes cyber a powerful tool, a dangerous weapon, and an unbridled and sometimes unpredictable IRC. With this image in mind, we must understand the power of such a tool and how, as it ties the other IRCs together through networks, communications, and other technological and infrastructure enablement, cyber is also a delicate and powerful tool.

Cyber, especially in the US collection of IRCs is a capability unmatched by any other power in the world. “US skills at cyberwar have no equal. US institutions lead the world in the commercialized arts of persuasion, and the collection and analysis of personal information for commercial and political purposes have proceeded farther in the United States than anywhere else. No country is more advanced in digitizing and networking things” (Libicki, 2017). This is also relevant in relation to cyber capability across the spectrum of not only warfare, but industry, banking and other important Critical Infrastructure (CI) umbrellas. “The use of cyber assets has been a form of force projection that helps initiate crises far ahead of and beyond the frontlines, creating forms of more complex crises that affect energy infrastructure, banking systems, and political leadership, and not solely the armed forces fighting on the frontlines. Again, the extension of traditional military conflict is not a new strategy, but new technologies have been able to provide both the means and vulnerabilities to allow such operations at a scale not often witnessed before, and with a smaller investment in resources on the part of the aggressor” (Danyk, et. al., 2017). It is evident that cyber has found its way into basically every area of life and it shows no sign of stopping. This is also evident in the fact that cyber has been established as a domain, specific to its own

capabilities and effects within the greater military construct. “The allocation of ‘domain’ status to cyberspace (alongside maritime, land, air, and space) serves a bureaucratic purpose to ensure that cyber operations (CO) receives sufficient financial and material support” (Argles, 2018). Overall, cyber has grown exponentially within its own sphere, replicating itself into the fissures of practically all other areas of military strategy, operations, and TTPs. Within the superstructure of IW, cyber holds a special place, encompassing the production of effects regardless of the IRC combination (see Figure 3). The cognitive affect from such rapid growth has been enormous with cyber becoming not only a term at the tip of every tongue, but a capability of which every entity desires a part. “Few cyber phenomena have captured the fascination of the media and the general public more than information theft through cyber exploitation and data exfiltration” (Jabbour & Devendorf, 2017). The terror and splendor inflicted on the collective considerations of the public show just how powerful and mature cyber has become and just how much we still have to learn.



**Figure 3:** Combined IW effects

As an IRC, ISR is one of the oldest with roots in warfare back to the dawn of recorded time. However, with cyber capabilities introduced in the 20<sup>th</sup> and 21<sup>st</sup> centuries, especially within the last two decades, ISR has become even more capable and powerful. As a discipline, there has never seemed to be any question concerning the power and utmost necessity of ISR. This is evident in the amounts of money invested in this IRC from the highest echelons of government with organizations such as the NSA, CIA, and FBI, all of which depend on ISR operability and capability to function. The great enabler in much of the maturation of ISR has been technology, again an area of obvious importance from the top down. With technology comes the need and desire to integrate other IRCs, most notably cyber capabilities, into the ISR capability framework. With this integration has come a new way of conducting ISR operations including the kinds of information sought and the kinds of information environments accessed and used. After the breakdown of IW in the 90s, ISR and the other silos of IRCs continued on parallel paths. “The ISR community kept building and operating systems of greater acuity and range. Electronic warriors went back to mastering their magic in support of air operations, counter-improvised explosive devices, and other combat specialties. Psychological operators continued to refine the arts of persuasion and apply them to an increasing roster of disparate groups. Cyber warriors bounced through the space community before getting their own sub-unified command within which they could practice their craft” (Libicki, 2017). These parallel paths have characterized the ways in which ISR has expanded its own sphere of operational influence and continued to add to this important and versatile IRC. “A key component of such independent operability in both ISR and combat operations is the development and use of unmanned drones. The increasing use of drones for different functional areas (intelligence, electronic countermeasures, direct strikes, etc.) and different operational environments (land, sea, air, amphibious) is an important consideration for flexibility in dynamic conflict situations” (Danyk, et. al., 2017). With key capabilities like drone and other network-dependent operations has come the inescapable tie-in of cyber which has only served to abut these two fields even more closely. The recent merger of the Cyber 24<sup>th</sup> Numbered Air Force (NAF) and the ISR 25<sup>th</sup>

NAF into a new 16<sup>th</sup> NAF, makes the objective clear; a combined capability bringing with it not only cyber and ISR, but other IRCs as well.

ISR as a capability is also growing across the globe. “Foreign intelligence services use cyber tools in information-gathering and espionage. Several nations are aggressively working to develop information warfare doctrine, programs, and capabilities to enable a single entity to have a significant and serious impact by disrupting the supply, communications, and economic infrastructures that support military power” (Jabbour & Devendorf, 2017). With this in mind, it is important to see the advantages of such constructs and how NATO and the U.S. are going to meet the challenges of other nation states and the capabilities they continue to develop. The continued development of ISR as a capability has kept pace with and now has even combined with cyber, leading to a continued technology and IRC arc that shows every sign of culminating in a combined IW construct.

As a shift and evolution of cyber and ISR capabilities has occurred, EW has followed a similar trajectory. As technology and cyber and ISR capabilities progress, EW as an IRC finds itself at a distinct advantage due to the peculiar niche it fills. EW is focused on controlling, disabling, and manipulating various signals and devices from and within multiple electronic environments. “[E]lectro[magnetic] warfare can ... be carried out by controlling devices that emit radio-frequency (RF) energy. New forms of RF signals pervade homes and cities: Bluetooth, Wi-Fi, 5G, keyless entry systems, and Global Positioning System (GPS), to name a few. The coming Internet of Things (IoT) is essentially an Internet of RF-connected items. If software-defined radios (those capable of broadcasting or receiving signals over an arbitrarily selected frequency) become ubiquitous, they could be hijacked to jam or spoof targets hitherto inaccessible using traditional EW boxes” (Libicki, 2017). With this powerful reach into the RF spectrum, EW stands as an excellent, cyber enabled resource, capable of combining with other IRCs in many, powerful ways. Other nations such as China have recognized this powerful combination of capabilities for some time. “A 2004 White Paper on National Defense increased the PLA focus on “informationalization” and advocated the use of cyber and electro[magnetic] warfare in the early stages of a conflict” (Jabbour & Devendorf, 2017). Under these circumstances and with a full understanding of the scope of these capabilities, it is in the distinct interest of NATO and the U.S. to hone their own capabilities in this realm while leveraging the full power of other IRCs. Again, Russia is already moving forward with this philosophy: “Russia has ... developed multiple capabilities for information warfare, such as computer network operations, electro[magnetic] warfare, psychological operations, deception activities, and the weaponization of social media, to enhance its influence campaigns” (Majir & Vailliant, 2018). China has pronounced provenance as well as is related to EW: “In early writings, Major General Dai Qingmin stated, ‘the destruction and control of the enemy’s information infrastructure and strategic life blood, selecting key enemy targets, and launching effective network-electro[magnetic] attacks.’ He argued that this integration of cyber and electronic warfare would be superior to the US military’s approach at the time of network-centric warfare” (Kania & Costello, 2018).

EW is another IRC that has persisted for much of the 20<sup>th</sup> and 21<sup>st</sup> centuries. However, there has been a marked growth in capability with the advent of cyber and the continuing growth and expansion of ISR and IO that have led to a closer tracking of these capabilities, now seen from a holistic perspective. As these IRCs continue to cross streams and implement the others’ precious proficiencies, the need for closer attention and support from NATO and the U.S. will be necessary.

Information Operations like ISR has been used in war for literally thousands of years. However, IO looks at information in a way distinct from the other IRCs, especially as it relates to psychological influence and the power of propaganda. “[T]he NATO Allied Joint Doctrine for Psychological Operations states that Information Operations are defined as ‘a staff function that analyzes, plans, assesses and integrates information activities to create desired effects on the will, understanding and capability of adversaries, potential adversaries, and North Atlantic Council (NAC) approved audiences in support of Alliance mission objectives’” (Bialy, 2017). With the creation and proliferation of social media, IO has become an extremely powerful tool in the world of cyber and ISR specifically. IO also draws power significantly from cyber as an enabling force. IO has been used for centuries as a way to influence, deter, and coerce through non-kinetic and generally non-lethal means. “Nonlethality and ambiguity, for their part, may be exploited to modulate the risk of reprisals—notably, violent reprisals—for having carried out information operations” (Libicki, 2017). This technique combined with other, non-lethal means such as cyber and EW can generate power across the battlespace at many levels. China has used such integration and should be expected to continue this strategy into future conflicts in peace and in war. “The SSF’s cyber corps approach the cyber domain in a much more comprehensive way, reflecting a highly integrated approach to information operations that actualizes critical concepts from PLA strategic and doctrinal



approaches” (Kania & Costello, 2018). Other nations recognize the flexibility and power of IO as well as other advantages, including scalability, portability, cost, and ambiguity. “Russia recognizes that information operations offers an opportunity to achieve a level of dominance... it provides a significantly less costly method of conducting operations since it replaces the need for conventional military forces” (Majir & Vaillant, 2018). It is difficult not to see how powerful IO is in regards to influence and dominance since information has become and remains a key to everything from business to commerce to military operations, especially as it relates to social media. “[A]part from its monetizing potential, social media has also become an excellent channel to mobilize support, disseminate narratives, wage information operations, or even coordinate military operations in the real world. States and non-state actors have started to extensively use social media to influence perception, beliefs, opinions and behaviors of their target audiences” (Bialy, 2017). The mature capability of IO across the globe and in and through organizational constructs lends itself well to the growth potential of IW, making it an undeniable asset in the combined scope of IW capabilities.

The mature capabilities manifested in and through cyber, ISR, EW, and IO respectively tend to culminate in a combined IW merger that can harness and exploit all of these competencies in myriad combinations. “[G]iven today’s circumstances, in contrast to those that existed when information warfare was first mooted, the various elements of IW should now increasingly be considered elements of a larger whole rather than separate specialties that individually support kinetic military operations” (Libicki, 2017). This concept continues to build as U.S. military services continue to not only merge capabilities into unified commands, but even characterize those commands as IW-centric. This philosophy of IRC warfare prosecution follows the emerging and established trends of the Russian and Chinese military complexes while further combining critical IRCs in a manner that will enable current and future warfare for decades to come.

#### 4. JADC2 and IW: Enabling and driving joint all-domain operations

The fact that JADC2 and IW are both strategies growing in parallel and power makes right now a perfect time to interleave these concepts. Both scaffolds include multiple capabilities, integration of technology and communications, and combined battlespace effects making the ultimate confluence of the two intuitively practical. However, it is a given that this is no easy task; one fraught with extreme complexity, culture clash, and opportunity for miscommunication. These difficulties, however, only highlight the pronounced need to get underway with their integration and get out far ahead of our adversaries. “It’s a bold approach, one that takes what the US military calls jointness to a new level” (Clark, 2020). This paradigm shift includes multiple levels of combinatory power and effects (see Figure 4). The concept of IW and JADC2 for JADO includes overlaying and matrixing the strategies, capabilities, and effects in myriad fashion.

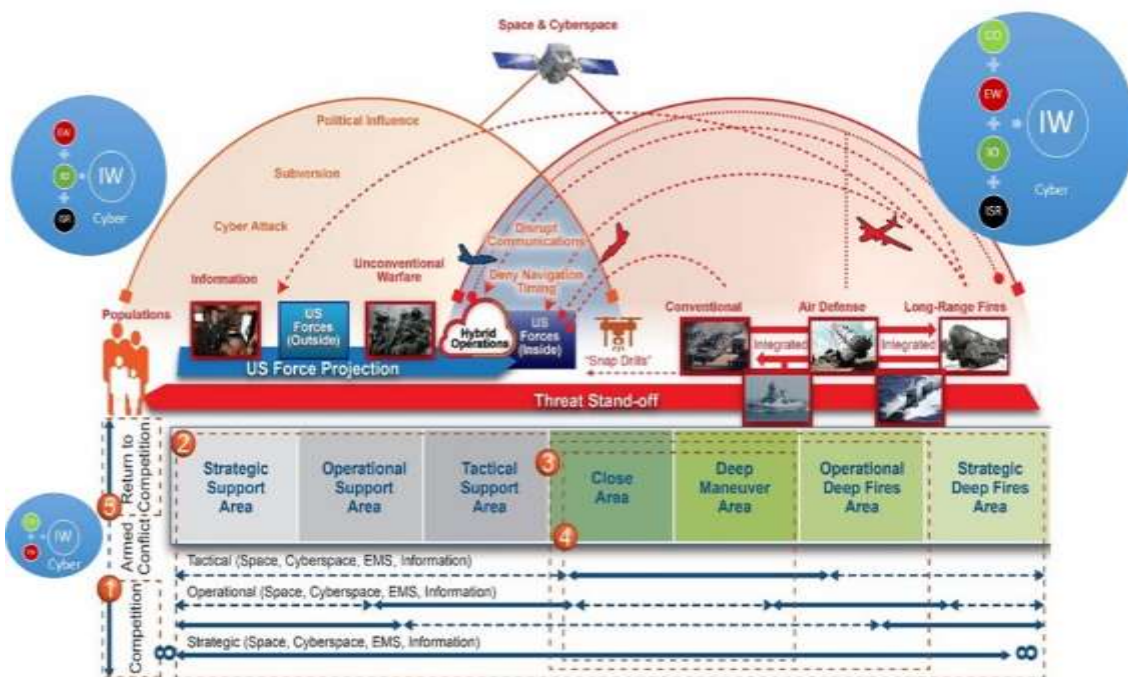
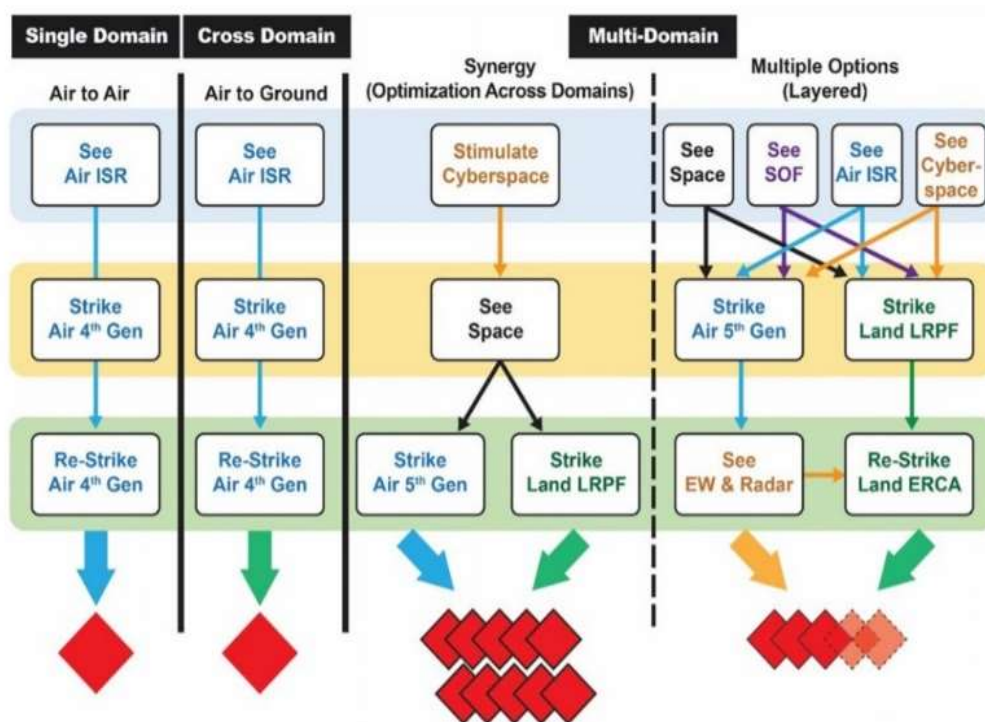


Figure 4: IW enabled JADC2

In order to bring these concepts closer together, further ties will have to be created and reoriented between combatant commands and other structures. “For example, the commanders of Space Command and Cyber Command have authority over operations in the space and cyberspace domains. All-domain operations, however, will integrate space and cyberspace with operations in the air, land, and sea domains, where geographic commanders traditionally have authority. Therefore, to integrate all domains—including space and cyberspace—JADC2 will require new links between combatant commands” (Dwyer, 2020). This is another area where IW will need to be woven into the JADO concept. With cyber as an IRC along with IRS, IO, and EW, the combinatory power of these capabilities must be interspersed with JADC2 in order to ensure JADO is made possible. Without a strong tie to the information piece of warfare, JADO cannot become a reality.

Information is not just words, pictures, video, and other bits of communicative matter, but rich, processed data such as that used in IO, ISR, EW, and CO. “To the three traditional domains of warfare – land, sea and air – space and cyber have been added. Some strategists include information as its own domain” (Goure, 2019). The central importance of information for strategic and operational penetration and supremacy is the crux of IW. Through the use of IW, JADC2 is tied together and enabled to fully leverage all domains for JADO. The effects made possible through the use of IW and JADC2 for JADO can be seen in Figure 5. (Kasubaski, 2018) Combined JADO and IW Effects.



**Figure 5:** Combined JADO and IW effects

The growth and potential of JADC2 and IW together to enable JADO is recognized by the joint force and continues to permeate all areas of modernization and strategic development. “The Army leadership recognizes that changes in the character of warfare will take place, and that these changes are unpredictable. The goal of the modernization strategy is to set the conditions for the Army to adapt to those changes better than any possible rival” (Wille, 2019). This push combined with the previously mentioned Air Force and joint service commitment to establishing JADO makes the integration of JADC2 and IW all the more important. Nakasone and Lewis give a direct look into this possibility as they relate JADO to the IRC, IW combination of EW, CO, and IO: “The harnessing of the electromagnetic spectrum and the advent of modern communications technologies have allowed militaries with advanced warfighting capabilities to seize the advantage by engaging in multiple domain battle” (Nakasone & Lewis, 2017). This multi-level technologically potent combination across the all-domain spectrum makes for a potential game changing battlespace in which joint operations using IW will continuously overpower adversary capabilities.

### 5. JADC2 and IW for JADO conceptual model

The following is an explanation of the JADC2 and IW for JADO Conceptual Model (Figure 6) (Sipper, 2021). The model is built from the other referenced figures in this paper to depict how the strategic concepts can be flowed together in a way to create combined IW and JADC2 exponential effects. Rolfe, et. al., characterize the complexities of multi-domain conflict well when they state, “There are significant challenges in understanding a situation. First, there is a large amount of data relevant to a situation and it changes constantly. The topology of nodes, links, nodal equipment, architecture, protocols, and networks is always in flux. Also, network traffic is changing, together with software applications for the user and for the managers of the networks” (2014). The JADC2 and IW for JADO Conceptual Model is an effort to provide a high-level strategic view of these complexities and interconnections on which to further build operational capabilities and concepts.

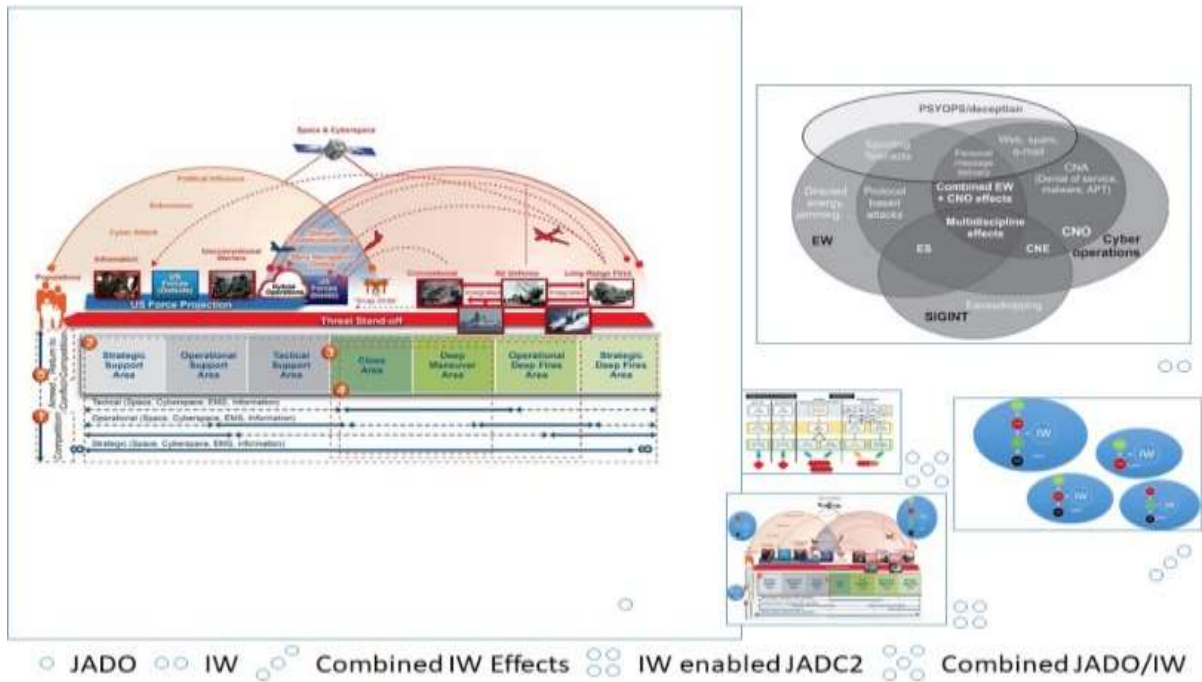


Figure 6: JADC2 and IW for JADO conceptual model

The model begins with the overarching concept of JADO, recognizing that all domains must first be interlaced and integrated such that operations from a joint perspective are possible. The next level of the model includes a graphic representation of the IW IRCs, indicating how these capabilities overlap and interoperate to produce effects. The third level graphic explains how IW occurs at multiple levels. The concept is not an all-or-nothing proposition, but one of combined capabilities from many different IRC perspectives. However, regardless of the combination of IRCs, cyber is ubiquitous, providing the enabling domain capability for all of the associated IW components. IW enabled JADC2 basically takes the Combined IW Effects and overlays them onto the JADO concept to show how IW can be integrated to enable JADC2 for JADO. Finally, Combined JADO and IW Effects are presented in the fifth and last level of the model indicating the culmination of the JADC2/IW combinatory modalities. The conceptual model is ultimately a way to see how all of these strategic and operational concepts can be streamed in such a way to outstrip adversary strategy and power.

### 6. Conclusion

The collective supremacy of JADC2 and IW is a case involving a great deal of complex and ramified concepts. JADO is an area of recent, prolific, and profound expansion and will likely continue with theory, doctrine, and joint operations and exercises characterizing its definitive establishment. IW is growing and becoming ingrained in parallel with JADO, making the time for integration of these super strategies suitable. With the integration of JADC2 and IW for JADO, the relationships between IRCs and all-domain interoperability stands to benefit greatly and proliferate into strategy and operations prodigiously. The JADC2 and IW for JADO Conceptual Model is one way to look at these relationships and obtain a strategic and operational view of this potentially game changing combination. Through the use of IW, JADO can not only be made to work better, but to project joint force power deeper into all battlespaces now and far into the future.

## References

- Argles, C. (2018). A Conceptual Review of Cyber-Operations for the Royal Navy, *The Cyber Defense Review*, Vol. 3, No. 3 (FALL 2018), pp. 43-56.
- Bialy, B. (2017). Social Media—From Social Exchange to Battlefield, *The Cyber Defense Review*, Vol. 2, No. 2 (SUMMER 2017), pp. 69-90.
- Clark, C. (2020). Gen. Hyten On The New American Way of War: All-Domain Operations, *Breaking Defense*, Accessed 4/26/2020: <https://breakingdefense.com/2020/02/gen-hyten-on-the-new-american-way-of-war-all-domain-operations/>
- Danyk, Y. Maliarchuk, T. and Briggs, C. (2017). Hybrid War: High-tech, Information and Cyber Conflicts, *Connections*, Vol. 16, No. 2 (Spring 2017), pp. 5-24.
- Dwyer, M. (2020). Making the Most of the Air Force's Investment in Joint All Domain Command and Control, *Center for Strategic and International Studies*, Accessed 4/26/2020: <https://www.csis.org/analysis/making-most-air-forces-investment-joint-all-domain-command-and-control>
- Goldfein, D. (2018). Enhancing Multi-Domain Command and Control: Tying it All Together, Air Force Chief of Staff Policy Letter, Accessed 4/26/2020: [https://www.af.mil/Portals/1/documents/csaf/letter3/Enhancing\\_Multi-domain\\_CommandControl.pdf](https://www.af.mil/Portals/1/documents/csaf/letter3/Enhancing_Multi-domain_CommandControl.pdf)
- Goure, D. (2019). A New Joint Doctrine for an Era of Multi-Domain Operations, *RealClear Defense*, Accessed 4/26/2020: <https://www.tradoc.army.mil/Publications-and-Resources/Article-Display/Article/1987883/a-new-joint-doctrine-for-an-era-of-multi-domain-operations/>
- Jabbour, K. and Devendorf, E. (2017). Cyber Threat Characterization, *The Cyber Defense Review*, Vol. 2, No. 3 (FALL 2017), pp. 79-94.
- Kania, E. and Costello, J. (2018). The Strategic Support Force and the Future of Chinese Information Operations, *The Cyber Defense Review*, Vol. 3, No. 1 (SPRING 2018), pp. 105-122.
- Kasubaski, B. (2018). Exploring the Foundation of Multi-Domain Operations, *Small Wars Journal*, Accessed 5/4/2020: <https://smallwarsjournal.com/jrnl/art/exploring-foundation-multi-domain-operations>
- Libicki, M. (2017). The Convergence of Information Warfare, *Strategic Studies Quarterly*, Vol. 11, No. 1 (SPRING 2017), pp. 49-65.
- Majir, M. and Vaillant, B. (2018). Russian Information Warfare: Implications for Deterrence Theory, *Strategic Studies Quarterly*, Vol. 12, No. 3 (FALL 2018), pp. 70-89.
- Nakasone, P. and Lewis, C. (2017). Cyberspace in Multi-Domain Battle, *The Cyber Defense Review*, Vol. 2, No. 1 (WINTER 2017), pp. 15-26.
- Orye, E. and Maennel, O. (2019). Recommendations for Enhancing the Results of Cyber Effects, 2019 11th International Conference on Cyber Conflict
- Rolfe, R., Louisa-Lamos, F., Odell, L., Agre, J., Gordon, K., Alspector, A., and Barth, T. (2014). 19th ICCRTS Cyber Operations Model for Multi-Domain Conflict, Institute for Defense Analyses.
- TRADOC PAM 525-3-1 \*2018, The US Army in Multi-Domain Operation 2028.
- Underwood, K. (2020). U.S. Army Sets Aside Money for Joint All-Domain Operations, SIGNAL AFCEA, Accessed 4/26/2020: <https://www.afcea.org/content/us-army-sets-aside-money-joint-all-domain-operations>
- Wille, D. (2019). A Summary of Multi-Domain Operations, *New America*.



# Defensive Cyber Deception: A Game Theoretic Approach

Abderrahmane Sokri

DRDC CORA, Ottawa, Canada

[Sokriab@gmail.com](mailto:Sokriab@gmail.com)

DOI: 10.34190/EWS.21.077

**Abstract:** While traditional protective and reactive measures in cyberspace are crucial for cyber security, they cannot be a panacea against all sophisticated and well organized adversaries. Cyber deception reasoning has been recognized as a well-suited solution to enhance traditional security controls. Deception is a technique used to mislead attackers, increase their uncertainty, and push them to behave against their interests. This paper offers a new game formulation and a formal discussion on the strategic use of honeypots in network security. The adversarial interaction is formulated as a leader-follower game where the defender disguises honeypots as normal systems. Results indicate that the attacker would target the most valued system no matter what its state is (fake or normal). The most valuable target is identified using the financial concept of Exceedance Curve. This curve is derived by randomizing each reward and cost in the expected utilities of the defender and the attacker. A case study is presented and discussed to illustrate the suggested game and characterize its equilibria.

**Keywords:** game theory, leader-follower game, problem of common knowledge, cyber-defence, cyber-attack, cyber-security, exceedance curve, deception, honeypot

## 1. Introduction

There is an ongoing race between attackers and defenders in cyberspace. As soon as a new solution is found, a more sophisticated technique to bypass it is established. Traditional defensive measures that prevent future cyber-attacks are still crucial mechanisms in cyberspace, but they are not a panacea (Roy et al. 2010). In addition to cryptography and tamperproof techniques, these protective measures include a number of practices as summarized in Table 1 (Sokri, 2020).

**Table 1:** Traditional detection/ prevention measures in cyberspace

Technique	Definition
Antivirus programs	Software used to scan devices, detect signs of malware presence, and remove them.
Firewalls	Security controls used to limit access to private networks connected to the Internet.
Intrusion Detection Systems (IDS)	Algorithms used to detect suspected intrusions and alert the security specialists in real-time.

Defenders can also use deception as a defence strategy to enhance the effectiveness of their security. Deception is an active manipulation of reality (Almeshekah and Spafford, 2016) aiming to mislead attackers, increase their uncertainty, and push them to behave against their interests (Carroll and Grosu, 2001; Zhu, 2019). Throughout the history, deception has been widely seen as a force multiplier in war and military conflicts. Its foundations date back to ancient civilizations. Since the seminal book by the Chinese military strategist Sun Tzu, every military admits that deception plays a prominent role in any warfare. When the friendly forces (Blue Team) are able to attack, they must seem unable; when they are active, they must seem inactive; when they are near, they must make the adversary (Red Team) believe they are far away; when far away, they must make the enemy believe they are near (Tzu, 2012; Cohen, 1998; Heckman et al., 2015; Almeshekah, 2015).

Deceptive techniques can be divided into two main acts: (1) dissimulation and (2) simulation (Bell and Whaley, 1991; Almeshekah, 2015). Dissimulation hides the existence, the nature, or the real value of targets. Simulation displays false information. As shown in Table 2, each act consists of three basic components where the deceiver manipulates reality by hiding it, altering it, or manufacturing it.

**Table 2:** Simulation and dissimulation techniques

Dissimulation		Simulation	
Components	Definition	Components	Definition
Masking	Making the real undetected	Mimicking	Making the real to look like something else
Repackaging	Making things appear different	Inventing	Creating a new reality
Dazzling	Confusing the target with others	Decoying	Driving attention away

As the evolutionary advancement in information and communication technologies pervades into business transactions and military operations, deception finds new fertile ground (Chou and Zhou, 2012). Successful

deception in cyberspace depends on the information asymmetry between the deceiver and the deceivee (Zhu, 2019). The deceiver can employ a honeypot as a normal system to increase the deceivee’s uncertainty and effort to determine whether the system is true or fake (Cohen, 1998; Rowe et al., 2007; Carroll and Grosu, 2011, Sokri, 2020).

Honeypots are totally independent systems with no valuable information. They give service providers the ability to disconnect and analyze the system after a given attack without any unwanted interruptions. As shown in Table 3, honeypot-based tools are generally designed to accomplish four main missions (Almeshekah, 2015): The detection, prevention, research and response to cyber-attacks.

**Table 3:** Security applications of Honeypots

Mission	Strategic objectives
Detection	Detect and stop spams
Prevention	Slow down attacks Confuse attackers Transfer risk to the attacker’s side Deter attacks
Research	Investigate the latest attacks Analyse new families of malware
Response	Preserve the attacked system’s state Simplify the forensic work

Game theory has been used as a sound theoretical foundation for understanding the information asymmetry between deceivers and deceivees in network security (Baston and Bostock, 1988). The analytical setting can be non-cooperative (e.g., a game between defenders and attackers) or cooperative (e.g., sharing collective costs or rewards between allies). It can be discrete or continuous, static or dynamic, deterministic or stochastic, and linear or non-linear (Sokri, 2020). Many researchers including Rowe et al. (2007), Garg and Grosu (2007), Carroll and Grosu (2011), and Almeshekah (2015) have recognized the game theoretic reasoning as well-suited to defensive cyber deception.

The aim of this paper is to show how game theory can guide defensive deception against cyber-attacks. It offers a new game formulation and a formal discussion on the strategic use of honeypots in network security. The paper is organized into five sections. Following the introduction, we set up a new deception game in section 2. In section 3, we provide the deception game’s equilibria. A numerical illustration and a formal discussion are provided in section 4. Some concluding remarks are indicated in section 5.

**2. The deception game**

Consider a security game between an attacker *a* (the follower) and a defender *d* (the leader) in a computer network. Let  $A = \{s_1, s_2, \dots, s_n\}$  be a set of *n* systems that the attacker may choose to attack. Following Rowe et al. (2007) and Carroll and Grosu (2011), we assume that the defender seeks to prevent attacks by disguising  $m \leq n$  honeypots as normal systems. The purpose of this camouflage is to detect unauthorized access and record their methods of attack. In this stochastic framework, each player is faced with a binary decision with Bernoulli distribution. Let  $X_i$  be the random variable taking the value 1 with probability  $\delta_i$ , if the system *i* is a honeypot and the value 0 with probability  $1 - \delta_i$ , if the system is normal. Similarly, denote by  $Y_i$  another random variable taking value 1 with probability  $\rho_i$ , if the system *i* is attacked and the value 0 with probability  $1 - \rho_i$ , otherwise.

Let  $U_d^h(s_i)$  be the defender’s payoff if the attacked system  $s_i$  is a honeypot and  $U_d^n(s_i)$  her payoff if the system is normal. Similarly, denote by  $U_a^h(s_i)$  the attacker’s payoff if the attacked system  $s_i$  is a honeypot and by  $U_a^n(s_i)$  the attacker’s payoff if the attacked system  $s_i$  is normal. The expected utilities of the defender and the attacker are respectively given by the following cost-benefit formulations

$$E(U_d) = \sum_{i=1}^n \rho_i (\delta_i U_d^h(s_i) - (1 - \delta_i) U_d^n(s_i)) \tag{1}$$

$$E(U_a) = \sum_{i=1}^n \rho_i ((1 - \delta_i) U_a^n(s_i) - \delta_i U_a^h(s_i)) \tag{2}$$

The expected utilities in equations 1 and 2 depend simply on the attacked systems and their state (fake or normal).

Fixing the value of  $\delta_i$ , the first problem to solve is to find the attacker's best response to  $\delta_i$ . This optimization problem can be formulated as a linear program where the attacker maximizes her expected utility given  $\delta_i$ .

$$\text{Max } \sum_{i=1}^n \rho_i \left( (1 - \delta_i) U_a^n(s_i) - \delta_i U_a^h(s_i) \right) \quad (3)$$

$$\text{subject to } \sum_{i=1}^n \rho_i = 1 \quad (4)$$

$$\rho_i \geq 0, \forall i \quad (5)$$

The two constraints in equations 4 and 5 define the set of feasible solutions  $\rho$ . To simplify the problem, we assume that the attacker will play only pure strategies. A pure strategy means that the attacker will choose to attack a single target.

Denoting by  $\rho_i(\delta_i)$  the follower's best response to  $\delta_i$ , the defender seeks to solve the following problem:

$$\text{Max } \sum_{i=1}^n \rho_i(\delta_i) \left( \delta_i U_d^h(s_i) - (1 - \delta_i) U_d^n(s_i) \right) \quad (6)$$

$$\text{subject to } \sum_{i=1}^n \delta_i = m \quad (7)$$

$$\delta_i \in [0,1], \quad \forall i \quad (8)$$

The defender is assumed to adopt a mixed strategy. In this case, the defender can assign a probability distribution over the set of targets in order to avoid being predictable. A mixed strategy will consist of a vector of pure strategies  $\delta$  (Walker and Wooders, 2006; Jain et al., 2010; Coniglio, 2013). The two constraints in equations 7 and 8 enforce the leader's mixed strategy to be feasible.

It is often assumed in security games that the players know their own payoffs and the payoffs of their opponents. This assumptions is not always true in most real-world cyber-security problems and can make the committed strategies ineffective (Coniglio, 2013; Sokri, 2018). This issue is known as the problem of common knowledge in cyberspace. To solve this problem, we randomize each reward and cost using stochastic simulation. Instead of using single deterministic values, we change the static value of each variable to a range of values using three-point estimates (optimistic, most likely, and pessimistic).

### 3. Deception game's equilibria

In the concept of leader-follower equilibrium, also known as Stackelberg equilibrium, the leader anticipates the attacker's reaction and credibly determines and commits to a certain strategy. The best reaction of the follower at the equilibrium will maximize the leader's payoff (Coniglio, 2013). Assume that the attacker will attack a single target, it is clear to note that the optimal strategy for the follower is to attack the system  $s_k$  ( $\rho_k = 1$ ) that maximizes her expected payoff  $\left( (1 - \delta_k) U_a^n(s_k) - \delta_k U_a^h(s_k) \right)$ . This means that the attacker would target the most valued system no matter what its state is (fake or normal). The defender would seek to prevent this attack by employing a honeypot.

To identify the target that serves best the follower, we will use the financial concept of Exceedance Curve (Hubbard and Seiersen, 2016; Sokri, 2019). To do so, consider the following continuous random variable

$$U_k = \left( (1 - \delta_k) U_a^n(s_k) - \delta_k U_a^h(s_k) \right). \quad (9)$$

Let  $F$  be its Cumulative Distribution Function (CDF) and  $G$  its Complementary Cumulative Distribution Function (CCDF). As shown in equation 10, for each potential payoff  $u_k$ ,  $G(u_k)$  is the probability of obtaining a value greater than  $u_k$ .

$$G(u_k) = 1 - F(u_k) = P(U_k > u_k). \quad (10)$$

If, for example, the  $n^{\text{th}}$  system has the highest expected payoff, this defender-attacker Stackelberg game will have multiple equilibria of the form

$$\langle \delta = (\delta_1, \delta_2, \dots, \delta_{n-1}, 1), \rho = (0, 0, \dots, 0, 1) \rangle. \tag{11}$$

This standard solution is in par with the existing literature. It is particularly consistent with the studies conducted by Jain et al. (2010) and An et al. (2011) in the physical world. This literature suggests that attackers aim first the most appreciated target. While game theory is used as a common ground to inspire the development of defence algorithms in this literature, it is worthwhile to note two fundamental differences between our paper and this literature. Jain et al. (2010) and An et al. (2011) show how a game theory can be used to optimally allocate resources in the physical world. Jain et al. (2010), for example, applied game theory to assign air marshals to protect flights in airports. This paper shows how deception as a defence multiplier can be formulated in cyberspace where security is more complex than in physical domains.

**4. Numerical illustration**

To illustrate the suggested method, consider a computer network of four systems in which the defender seeks to disguise honeypots to prevent potential attacks. As shown in Table 4, the defender can use a fake target and get a reward, if the target is attacked. She can also leave the target normal and incur a cost, if it is attacked. The attacker can attack a target and get a reward if the target is normal. If the target is a honeypot, she will incur a cost. To solve the uncertain observability problem, a three-point estimate approach is used to assess the likely fluctuation of each rewards and costs in Table 4. In this table, only the most likely values are provided. Details on the optimistic and pessimistic values are omitted. They will, however, be used in the stochastic simulation.

If we know the minimum, the most likely, and the maximum values of a given uncertain variable, two common probability distributions are particularly suitable to assess its likely fluctuation: Triangular and Program Evaluation and Review Technique (PERT) distributions. Assuming asymmetry in the distributions of rewards and costs, we used a PERT distribution in this illustration (Sokri and Solomon, 2013; Sokri and Ghanmi, 2017). We assume in this leader-follower game that the expected value of the Bernoulli random variable  $X_i$  is  $\delta_i = 0.5$ , for each system  $s_i$ .

**Table 4:** Payoff table

	Defender		Attacker	
	Most likely reward	Most likely cost	Most likely reward	Most likely cost
System 1	5	3	11	3
System 2	1	1	5	4
System 3	5	3	11	7
System 4	3	2	8	5

For each system  $s_k$ , a PERT distribution and a stochastic simulation are used to randomize the corresponding potential payoff. The outputs from these simulations are used to derive the Exceedance Curves as shown in Figure 1. As mentioned before, these curves depict for each target the likelihood of exceeding a given value of payoff. The red curve, for example, shows that the probability of earning a payoff of 3.19 or more by attacking System 1 would be 97.5%. The other Exceedance Curves indicate that the probability of reaching this payoff by attacking the other systems would be approximately 0%. Moreover, Systems 2 and 4 may have negative payoffs since there are situations where their costs overweight their benefits. This results shows that the most valuable target for the attacker is System 1. The attacker will, therefore, choose to attack this target and the defender will choose to protect it. In this case, the interaction between the two players will have multiple equilibria of the form

$$\langle \delta = (1, \delta_2, \delta_3, \delta_4), \rho = (1, 0, 0, 0) \rangle. \tag{12}$$

A solution that satisfies the constraints can be formulated as

$$\langle \delta = (1, 0.5, 0.25, 0.25), \rho = (1, 0, 0, 0) \rangle. \tag{13}$$

A Pareto dominance analysis should be conducted to refine the equilibrium (An et al., 2011).

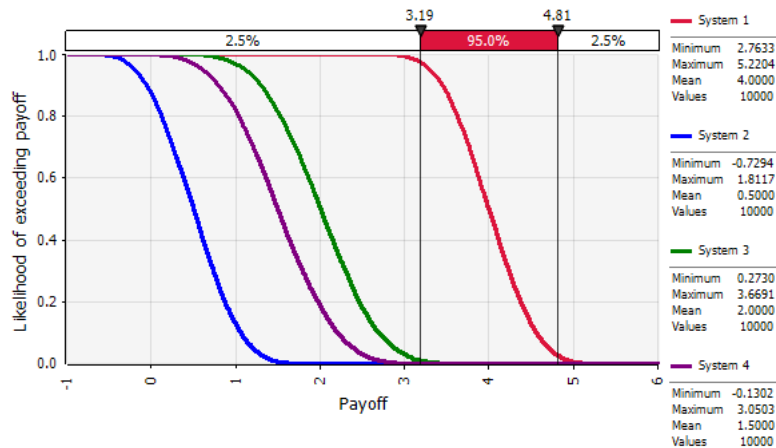


Figure 1: Exceedance curves

This standard solution indicates that the game-theoretic formalism can be adapted to obtain sound solutions in cyberspace. But it also highlights many challenges and raises many research questions associated with the applicability of game theory in the cyber world. In addition to the complexity and the dynamic nature of cyberspace, the application of game theory faces at least two main challenges in this domain (Sokri, 2020):

- It is often assumed that the security game is played between two known and rational adversarial agents. In real-world, the interaction generally involves multiple unknown attackers, with bounded rationality, and from multiple locations.
- Moreover, it is implicitly supposed that each player is able to investigate and process a large quantity of information in a methodical and accurate manner. In real-world, this assumptions is not always valid.

Using the game-theoretic formalism under these simplified assumptions can provide a sound theoretical foundation for understanding the strategic interactions in network security. But it can also result in some ineffective defensive strategies.

Scaling up the problem to a real-world-sized framework may make it intractable and much more complex to understand. It is, therefore, necessary to adopt an interdisciplinary approach in these cybersecurity problems. In addition to traditional measures in cyberspace such as antivirus software and firewalls (Shiva et al., 2012), many other techniques can be combined with game theory to analyse cyber conflicts. As stated by the United States Department of Defence (DoD, 2011), these techniques particularly include (but are not limited to): Computer Simulation, Genetic Algorithms, Graph Theory, Reliability Modelling, and Cyber Forensic Analysis.

## 5. Conclusion

This paper uses a game theoretic approach to show how deception as a force multiplier can be formulated in cyberspace. The adversarial interaction is formulated as a leader-follower game where the defender disguises honeypots to detect, prevent, and analyse cyber-attacks without any unwanted interruptions. Results showed that the adversarial interaction between the two players will have multiple equilibria. In each equilibrium, the aggressor would target the most valued system, no matter what its state is, and the defender will choose to protect it using a honeypot. This standard solution is in par with the existing literature in the physical security world even if the adopted approaches, the desired objectives, and the domains of work are totally different.

To obtain sound and effective solutions in a reasonable time, an interdisciplinary approach should be adopted in this area. Using techniques such as computer simulation and cyber forensic analysis under a solid game-theoretic setting would provide the possibility to investigate and address many standard cybersecurity issues. These issues include (but are not limited to):

- the problem of common knowledge,
- the problem of bounded rationality, and
- the problem of multiple unknown attackers.

## References

- Almeshekah, M. H. (2015). Using deception to enhance security: A Taxonomy, Model, and Novel Uses.
- Almeshekah, M. H., and Spafford, E. H. (2016). Cyber security deception. In *Cyber deception* (pp. 23-50). Springer, Cham.
- An, B., Tambe, M., Ordonez, F., Shieh, E., and Kiekintveld, C. (2011). Refinement of strong Stackelberg equilibria in security games. *Proceedings of the 25th Conference on Artificial Intelligence*, p.587–593.
- Baston, V.J. and Bostock, F.A. (1988). Deception Games. *International Journal of Game Theory*, Vol. 17 (2).
- Bell, J. B. and Whaley, B. (1991). *Cheating and Deception*. Transaction Publishers, New Brunswick.
- Carroll, T.E. and Grosu, D. (2011). A Game Theoretic Investigation of Deception in Network Security. *Security and Communication Networks*, Vol. 4 (10), p. 1162–1172.
- Chou, H. M., and Zhou, L. (2012). A game theory approach to deception strategy in computer mediated communication. In *2012 IEEE International Conference on Intelligence and Security Informatics* (pp. 7-11). IEEE.
- Cohen, F. (1998). A Note on the Role of Deception in Information Protection. *Computers and Security*, Vol. 17 (6).
- Coniglio, S. (2013). *Algorithms for Finding Leader-Follower Equilibrium with Multiple Followers*. Ph.D. Thesis, Politecnico di Milano.
- Garg, N. and Grosu, D. (2007). Deception in Honeynets: a Game-Theoretic Analysis. *Proc. of the 8th IEEE Information Assurance Workshop*.
- Heckman, K. E., Stech, F. J., Thomas, R. K., Schmoker, B., and Tsow, A. W. (2015). *Cyber denial, deception and counter deception*. Springer.
- Hubbard, D.W. and Seiersen, R. (2016) *How to Measure Anything in Cybersecurity Risk*, Wiley, New York.
- Jain M, Tsai J, Pita J, Kiekintveld C, Rath S, Ordone, F, and Tambe, M. (2010). Software assistants for randomized patrol planning for the LAX airport police and the Federal Air Marshals Service. *Interfaces*, Vol. 40(4).
- Rowe, N.C., Custy, E.J., and Duong. B.T. (2007). Defending Cyberspace with Fake Honeypots. *Journal of Computers*, Vol. 2 (2), p. 25–36.
- Roy, S., Ellis, C., Shiva, S., Dasgupta, D., Shandilya, V., and Wu, Q. (2010). A Survey of Game Theory as Applied to Network Security. *Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS)*.
- Shiva, S., Bedi, H., Simmons, C., Fisher, M., Dharam, R. (2012). A Holistic Game Inspired Defence Architecture. *International Conference on Data Engineering and Internet Technology*.
- Sokri, A. (2018). Optimal resource allocation in cyber-security: A game theoretic approach. *Procedia computer science*, 134, 283-288.
- Sokri, A. (2019). *Cyber Security Risk Modelling and Assessment: A Quantitative Approach*. *Proceedings of the 19th European Conference on Cyber Warfare and Security (ECCWS19)*, Coimbra University, Portugal, pp. 466-474.
- Sokri, A. (2020). *Game Theory and Cyber Defence*. In *Games in Management Science* (pp. 335-352). Springer, Cham.
- The United States Department of Defence (DoD) (2011). *Cyber Intelligence Preparation of the Environment (CIPE)*. Technical Task Order 11-0002, Version 1.
- Tzu, S. (2013). *The Art of War*. Orange Publishing.
- Walker, M., and Wooders, J. (2006). Mixed strategy equilibrium. Retrieved December, 19, 2006.
- Zhu, Q. (2019). Game theory for cyber deception: a tutorial. In *Proceedings of the 6th Annual Symposium on Hot Topics in the Science of Security* (pp. 1-3).

# Using Semantic-Web Technologies for Situation Assessments of Ethical Hacking High-Value Targets

Sanjana Suresh, Rachel Fisher, Radha Patole, Andrew Zeyher and Thomas Heverin  
Drexel University, USA

[ss5264@drexel.edu](mailto:ss5264@drexel.edu)

[rcf49@drexel.edu](mailto:rcf49@drexel.edu)

[rdp74@drexel.edu](mailto:rdp74@drexel.edu)

[az458@drexel.edu](mailto:az458@drexel.edu)

[th424@drexel.edu](mailto:th424@drexel.edu)

DOI: 10.34190/EWS.21.070

**Abstract:** Ethical hacking consists of scanning for targets, evaluating the targets, gaining access, maintaining access, and clearing tracks. The evaluation of targets represents a complex task due to the number of IP addresses, domain names, open ports, vulnerabilities, and exploits that must be examined. Ethical hackers synthesize data from various hacking tools to determine targets that are of high value and that are highly susceptible to cyber-attacks. These tasks represent situation assessment tasks. Previous research considers situation assessment tasks to be tasks that involve viewing an initial set of information about a problem and subsequently piecing together more information to solve the problem. Our research used semantic-web technologies, including ontologies, natural language processing (NLP), and semantic queries, to automate the situation assessment tasks conducted by ethical hackers when evaluating targets. More specifically, our research focused on automatically identifying education organizations that use industrial control system protocols which in turn have highly exploitable vulnerabilities and known exploits. We used semantic-web technologies to reduce an initial dataset of 126,636 potential targets to 155 distinct targets with these characteristics. Our research adds to previous research on situation assessment by showing how semantic-web technologies can be used to reduce the complexity of situation assessment tasks.

**Keywords:** ontology modeling, situation assessment, target evaluation

## 1. Introduction

Ethical hacking consists of several phases including identifying targets, evaluating targets, gaining access, maintaining access, and clearing tracks. Evaluating targets consists of finding information about organizations such as IP addresses, hardware names/versions, software names/versions, ports, protocols, and services (Morris and Gao, 2013). For evaluating targets, ethical hackers also research vulnerabilities connected to the aforementioned information as well as exploits that can be used on the vulnerabilities.

Target evaluation often starts with finding targets based on the type of organization (organization-type) and the types of assets an organization holds (asset-type). High-value organization-types include organizations that, if attacked, could result in severe consequences for individuals, organizations, and communities. Examples of high-value organization-types include organizations that are identified as “critical infrastructure” organizations. According to the U.S. Cybersecurity and Infrastructure Security Agency (CISA), there are 16 critical infrastructure sectors that are so vital that “...their incapacitation or destruction would have a debilitating effect on security, national economic security, national public health or safety, or any combination thereof” (CISA, 2020). Table 1 shows the 16 critical infrastructure sectors according to CISA.

**Table 1:** Critical infrastructure sectors According to the U.S. Cybersecurity and Infrastructure Security Agency (CISA)

Chemical Sector	Commercial Facilities Sector	Communications Sector	Critical Manufacturing Sector
Dams Sector	Defense Industrial Base Sector	Emergency Services Sector	Energy Sector
Financial Services Sector	Food and Agriculture Sector	Government Facilities Sector	Healthcare and Public Health Sector
Information Technology Sector	Nuclear Reactors, Materials, and Waste Sector	Transportation Systems Sector	Water and Wastewater Systems Sector

In addition to ethical hackers identifying high-value targets based on organization-types, ethical hackers also select targets based off of asset-types. High-value asset-types across organizations include industrial control

systems (ICS). ICS are systems that control the physical operation of entities such as electrical systems, building automation systems, physical security systems, ventilation systems, military weapon systems, ship systems and more. ICS are found across all types of sectors including education, government, manufacturing, military, commercial, and electrical power. An attack on ICS could cause wide-spread electrical outages, disrupt everyday work activities, cause organizations to stop production, cause physical harm to people, and compromise safety (Morris and Gao, 2013).

Furthermore, highly exploitable ICS devices are considered to be even more valuable targets for ethical hackers. These types of devices are considered to have easy-to-exploit vulnerabilities which in turn have publicly known exploits. Many ICS protocols were originally designed without security in mind, resulting in a high level of risk for ICS. Vulnerabilities of ICS protocols include the lack of integrity, confidentiality, availability, authentication, authorization, and encryption. There are many types of cyber-attacks that can exploit these vulnerabilities such as denial-of-service, data injection, and command injection attacks (Morris and Gao, 2013).

Based on this information, an individual can reason that highly exploitable ICS devices (high-value asset-types) found within critical infrastructure organizations (high-value organization-types) are prime targets for ethical hackers. For this research paper, our focal point consisted of education organizations which fall under the "Government Facilities Sector" critical infrastructure category (CISA, 2020). Education targets include kindergarten through grade 12 (K-12) schools, institutions of higher education, and business and trade schools.

K-12 remote learning has suffered from a multitude of cyber-attacks. According to the U.S. Federal Bureau of Investigation (FBI), CISA, and the Multi-State Information Sharing and Analysis Center (MS-ISAC), K-12 schools are increasingly targeted by malicious cyber-attackers, "leading to ransomware attacks, the theft of data, and the disruption of distance learning services" (FBI, CISA, and MS-ISAC, 2020). These cyber-attacks are expected to continue throughout the course of remote learning and cause grave issues for K-12 schools due to their limited resources. Alongside K-12 schools, colleges and universities have increasingly become targets because of their new technology development, innovation, and research activities (Rogers and Ashford, 2015). Our research aims to reduce the complexity of target evaluation, a situation assessment task, in the education sector. The results of our research can contribute to better protecting K-12 schools, colleges, and universities.

## **2. Background**

The goal of our research was to explore how we can automate the target evaluation phase of ethical hacking. In this section, we thread together research on situation assessment and ontologies in the cybersecurity domain.

### **2.1 Situation assessment**

Situation assessment represents an information-intensive task that involves finding an initial set of information about a problem, evaluating information, finding more information, and integrating new information into existing evidence (Ben-Bassat and Freedy, 1982; Gorodetsky et al., 2005). Situation assessment also involves clarifying a problem and developing a plan to solve that problem (Kirillov, 1994; Noble, 1993; Salas et al., 2010). Once situation assessment is complete, it leads to obtaining situation awareness (Endsley and Garland, 2000). The process of forming a situation assessment is called completing a "situation assessment task" (Noble et al., 1986). Ben-Bassat and Freedy (1982, p. 489) consider a situation assessment task to be a "puzzle building task", as problem solvers synthesize information to understand a situation.

Previous research has shown that decision makers complete situation assessment tasks based on mental schemas that have been developed over years of experience. Lipshitz and Shaul (1997, p. 295) define schema as "situation or domain specific cognitive structures that (a) direct external information search, (b) specify which available information will be attended to and which information will be ignored, (c) organize information in memory, (d) direct the retrieval of information from memory, and (e) become more differentiated as a function of experience." Schemas provide a way for people to determine what information needs to be found, how to find the information, and how to piece the information together to make decisions (Lipshitz and Shaul, 1997; Noble et al., 1986; Serfaty et al., 1997). According to Elliot (2005, p. 215), "A schema helps determine what we attend to, what we perceive, and what we remember and infer."

In terms of the target evaluation phase of ethical hacking, situation assessment tasks involve the use of various resources to piece together information to select exploitable targets. This information includes organization



names and types, IP addresses, open ports, protocols in use, software in use, vulnerability information, exploit information, and more. Even though ethical hacking and other cybersecurity domains, such as cyber forensics, consist of situation assessment tasks, only limited research on situation assessment in the cybersecurity domain exists.

Research has been conducted on determining how cyber defenders, those who protect networks against cyber-attacks, synthesize various pieces of information to detect and analyze cyber-attacks (Barford et al., 2010; D'Amico et al., 2005). Cyber defenders must sift through multiple sources, such as network logs, email logs, server logs, computer logs and more, to determine if a cyber-attack is taking place and to understand what an attack is doing. Cyber defenders also use online sources to find information. These sources include the National Vulnerability Database (NVD), VirusTotal, Wireshark, U.S. Computer Emergency Readiness Team (CERT), and more.

Cyber defenders then fuse the data that they collect from multiple sources to determine what steps they need to take to contain and eradicate cyber-attacks on their networks (D'Amico et al., 2005; Goodall et al., 2009). The deluge of data that must be reviewed can cause an information overload for cyber defenders (D'Amico et al., 2005; Tyworth et al., 2012; Yen et al., 2010).

Although some research using ontologies to model domains within cybersecurity exists, there is a lack of research on how semantic-web technologies can be used to aid situation assessment tasks in the ethical hacking domain.

## **2.2 Ontologies and cybersecurity**

Ontologies are used to specify a domain, its entities, and the relationships among those entities (World Wide Web Consortium, 2015). Ontologies often provide the foundation of artificial intelligence (AI) systems. As Gruber stated (1993, p. 1), "To support the sharing and reuse of formally represented knowledge among AI systems, it is useful to define the common vocabulary in which shared knowledge is represented. A specification of a representational vocabulary for a shared domain of discourse — definitions of classes, relations, functions, and other objects — is called an ontology." An ontology is also often described as a web of data or a "semantic-web."

Ontologies are primarily made up of classes (categories of objects), objects that fall in those classes, object properties that link classes (as well as objects) with each other, and data properties that contain metadata about the objects. Resource description framework (RDF) is used as a standardized framework to describe all these ontology concepts (World Wide Web Consortium, 2015). After a domain is modeled with an ontology, one can run queries over the ontology to find new relationships and to discover new knowledge.

Queries over ontologies can help make decisions and form plans. As Gruber stated (1993, p. 3), a query system over an ontology "...takes descriptions of objects, events, resources, and constraints, and produces plans that assign resources and times to objects and events." The queries take inputs from an ontology and produce outputs to allow planners to perform their tasks. In terms of ethical hacking, an ontology could be used to ingest reconnaissance and scanning information to identify high-value targets which will allow ethical hackers to complete their tasks. Various semantic-web technologies can be used to query or search across a web of data. SPARQL, a recursive acronym for SPARQL query language, is a standard query language used for ontologies.

Previous research has focused on using ontologies for various reasons in cybersecurity. Grant, van't Wout, and van Niekerk (2020) created an ontology to assist with intelligence, surveillance, target acquisition and reconnaissance. Falk (2016) showed how ontologies can be used to organize open-source threat intelligence sources in order to build a comprehensive view of a threat environment. Aviad, Weceel and Abroamowicz (2015) focused on using ontologies to link known attacks with information technology (IT) products to develop a cybersecurity body of knowledge. Their research aimed to provide threats based off of a given configuration. Other research has focused on expanding the cybersecurity body of knowledge (Takahashi and Kadobayashi, 2015), creating an ontology that links together log and forensics data (Balduccini, Kushner, and Speck, 2015), and developing a taxonomy of ICS with an ontology (Flowers, Smith, and Oltramari, 2016). There lacks research on using an ontology for target evaluation.

### 3. ICS target ontology

In this section, we describe the structure of our ontology, called the ICS Target Ontology, and methods used to fill the ontology with relevant data to help with target evaluation.

#### 3.1 Ontology structure

In an ontology, classes represent categories of objects. An example class in the ICS Target Ontology is “ICS Protocol.” Representative objects in this class include popular ICS protocols such as BACnet, Modbus, S7, and Ethernet/IP. Another example class is “IP Address” which is filled with specific IP addresses. Other classes in the ontology include “Organization”, “Port Number”, “Vulnerability”, and “Exploit.”

Object properties are used to define relationships between classes; objects properties are then specified (filled out) to show specific links between two named objects. An example of an object property in the ICS Target Ontology is *usesProtocol* which defines a relationship between the classes IP Address and ICS Protocol. We can say that IP Address *usesProtocol* ICS Protocol. A specification of this object property is 123.45.67.89 *usesProtocol* BACnet.

The key object properties in the ICS Target Ontology are shown below in Table 2. A domain is the “subject” of an object property while the Range is the object of the object property.

**Table 2:** Object properties found in the ICS target ontology

Domain	Object Property	Range
Organization	<i>usesIPAddress</i>	IPAddress
IPAddress	<i>usesPort</i>	PortNumber
PortNumber	<i>usesProtocol</i>	ICSProtocol
ICSProtocol	<i>hasVulnerability</i>	Vulnerability
Vulnerability	<i>isExploitedBy</i>	Exploit

Within an ontology, a data property is used to define metadata about an object. An example data property includes “OrganizationType”, and for this data property we can state values such as “Education”, “Government”, “Manufacturing”, and more. A specification of this data property is: the data property value for the OrganizationType associated with Drexel University is Education.

The data properties in the ICS Target Ontology, their associated classes, and their possible values are shown below in Table 3.

**Table 3:** Data properties found in ICS target ontology

Data Property Name	Class	Possible Values
AttackComplexity	Vulnerability	Low or High
AttackVector	Vulnerability	Physical, Local, Adjacent Network or Network
AvailabilityImpact	Vulnerability	None, Low or High
ConfidentialityImpact	Vulnerability	None, Low or High
CVSS_BaseScore	Vulnerability	Numeric Score (out of 10)
ExploitabilitySubscore	Vulnerability	Numeric Score (out of 3.9)
ImpactSubscore	Vulnerability	Numeric Score (out of 6.0)
IntegrityImpact	Vulnerability	None, Low or High
OrganizationTargetValue	Organization	Low, Medium or High
OrganizationType	Organization	Education, IT, Communications, Food, Government, and More
PrivilegesRequired	Vulnerability	None, Low or High
UserInteraction	Vulnerability	None or Required

The vulnerability data properties are drawn from the NVD Common Vulnerability Scoring System Calculator (CVSS). The OrganizationType data property was derived from the critical infrastructure categories shown in Table 1.

The “OrganizationTargetValue” data property was derived from Factor Analysis of Information Risk (FAIR) which represents a model of risk factors and the relationships between the factors (Freund and Jones, 2014). As a standard quantitative model, FAIR serves as a widely accepted measure of assessing risks and cyberthreats. FAIR is comprised of two major categories: threats and assets. In the education sector, if a threat caused a denial-of-

service, the threat could significantly disrupt education organizations from providing education services and protecting the safety of the populations they serve. Therefore, education organizations were deemed to have an OrganizationTargetValue of “high.”

### 3.2 Data processing

Our ontology was created using Protégé, an open-source ontology editor. The classes, object properties, and data properties created “slots” for data ingestion. To find potential education organizations that use ICS protocols, we created searches in Shodan, an open-source search engine for Internet-connected devices. The searches focused on popular ICS protocols including BACnet, Modbus, S7 and Ethernet/IP. These searches generated over 126,000 publicly available ICS targets which are shown in Table 4 below. The Shodan data were then exported to a comma-separated values (CSV) file for further data processing.

**Table 4:** ICS protocol search results from Shodan

ICS Protocol/Port	Description of ICS Protocol	Number of IP Addresses from Shodan for ICS Protocols
BACnet/47808 (UDP)	Building automation systems protocol	21,136
Modbus/502 (TCP)	Popular protocol for ICS	14,523
S7/102 (TCP)	Siemens ICS protocol for programmable logic controllers (PLCs)	36,039
Ethernet IP/44818 (UDP)	Industrial Ethernet network protocol	54,938
	Total	126,636

The exported Shodan CSV file consisted of various data fields including organization name, IP address, ports, domain names, and more. The “organization name” data field provided an opportunity to determine how to identify organizations such as education organizations.

Two researchers manually examined a set of 1,500 entries from the Shodan CSV file to identify education organizations. There was a 99% agreement rate between the two researchers’ identification of education organizations.

We then used the Natural Language Toolkit (NLTK), which contains Python libraries and programs for statistical NLP, to determine how to automate the identification of education organizations. NLTK was used on the initial dataset (1,500 entries) that was manually examined. The initial success rate for identifying education organizations was 98.1% overall. At first, NLTK incorrectly identified organizations that had “universal” in their names as education organizations. We tweaked the NLTK libraries and then found a success rate of 100% based on the initial 1,500 entries.

From there, we used NLTK to analyze key words from each education organization name from the initial data set (such as “university”, “school”, “college”, and these terms in multiple languages) and used the Python library called “Wordnet” to build a list of synonyms. The list of synonyms was then used in NLTK to identify additional organizations that were not included in the initial dataset. For example, in the initial dataset, there were no entries that had “education” in the organization names. Wordnet generated “education” as a synonym for “school” which then allowed NLTK to identify entities like “Philadelphia Education Network” as an education organization. As another example, “higher education” was generated as a synonym for the terms “university” and “college.”

With the Wordnet synonym list, we ran NLTK over the Shodan CSV file of 126,636 entries. NLTK identified 2,272 education targets in 2.12 seconds and created a value for “Education” for Organization Type in the CSV file. Two researchers manually reviewed 500 targets of the NLTK CSV file and found agreement across all entries.

We used a utility within Protégé called “Create Axioms from Excel Workbook” that provides methods for ingesting data from CSV files into an ontology and linking data from the CSV files to ontology classes, object properties and data properties. For example, common vulnerabilities and exposures (CVEs) and their data properties were already in the ICS Target Ontology. Protégé allowed us to link targets directly to the CVEs upon importing the Shodan data.

Although 2,272 targets represent a considerably smaller number than 126,636 targets, a list of 2,272 targets still produces a complex situation assessment task in determining which targets are highly exploitable. Semantic queries, described in the next section, provide a method for finding targets that are highly exploitable.

## 4. Target evaluation query and results

### 4.1 SPARQL

While ontologies are used to model a domain, semantic queries are used to provide answers to questions about a domain. SPARQL is a set of specifications that provide languages and protocols to query ontologies (World Wide Web Consortium, 2013). SPARQL can be used to design simple queries (such as listing data property values for a specific object) and complex queries (such as determining the shortest path, among multiple paths, from one object to another object that is 100 objects away) (Angle and Guitierrez, 2008).

Types of queries available in SPARQL include union, filters, value aggregation, nested queries and more. SPARQL queries produce outputs, often in the form of tables, and can include solution modifiers such as distinct, order, and limit modifiers. Also, "...the output of a SPARQL query can be of different types: yes/no queries, selections of values of the variables which match the patterns, construction of new triples from these values, and descriptions of resources" (Pérez, J., Arenas, M., & Gutierrez, 2006, p. 30). As a result, SPARQL queries can be used to aid in situation assessment tasks completed by ethical hackers such as finding high-value targets that have highly exploitable vulnerabilities over a specific port number.

### 4.2 SPARQL complex query for target evaluation

Various queries were developed to assist with the situation assessment task of finding high-value targets that use ICS protocols which in turn have highly exploitable vulnerabilities and known exploits. We examined several types of queries focused on various object properties and data properties. After testing several queries and reviewing the query outputs, we developed a complex query focused on the following parameters:

- Target with "OrganizationType" value of *Education*
- Target with "Target Value" value of *High*
- Target that uses an ICS protocol via the object property *usesProtocol*
- Vulnerability with "Attack Complexity" value of *Low* (easy to attack)
- Vulnerability with "Attack Vector" value of *Network* (remotely exploitable)
- Vulnerability with "Privileges Required" value of *None* (unauthorized users can attack)
- Vulnerability with "User Interaction" value of *None* (no user interaction required)
- Vulnerability with "Availability Impact" value of *High* (can fully shut down the resource)
- A vulnerability has a publicly known exploit via the object property *isExploitedBy*

The full SPARQL query syntax is provided below. This query selects distinct organizations, rather than showing all IP addresses associated with each organization. Each organization could have multiple IP addresses that use the targeted ICS protocols.

```
SELECT
DISTINCT ?Target ?TargetType ?TargetValue ?Protocol ?CVE ?AttackVector ?AttackComplexity ?PrivilegesRequired ?UserInteraction ?AvailabilityImpact ?Exploit
WHERE {
    ?Target ics:OrganizationTargetValue ?TargetValue .
        FILTER (?TargetValue = "High") .
    ?Target ics:OrganizationType ?TargetType .
        FILTER (?TargetType = "Education")
    ?Target ics:usesIPAddress ?IPAddress .
    ?IPAddress ics:usesPort ?Port .
    ?Port ics:usesProtocol ?Protocol .
    ?Protocol ics:hasVulnerability ?CVE .
    ?CVE ics:AttackComplexity ?AttackComplexity .
```

```
?CVE ics:AttackVector ?AttackVector .
?CVE ics:PrivilegesRequired ?PrivilegesRequired .
?CVE ics:UserInteraction ?UserInteraction .
?CVE ics:AvailabilityImpact ?AvailabilityImpact .
    FILTER(?AttackVector = "Network" && ?AttackComplexity = "Low" && ?PrivilegesRequired =
    "None" && ?UserInteraction = "None" && ?AvailabilityImpact = "High") .
?CVE ics:isExploitedBy ?Exploit .
}
```

This query, which was completed in 0.3 seconds, produced 155 unique education organizations with the parameters listed above. Table 5 shows example results from this query.

**Table 5:** Sample query results for high-value education targets that use ICS assets with highly exploitable vulnerabilities and known exploits

Target	Target Type	Target Value	Protocol	CVE	Attack Vector	Attack Complexity	Privileges Required	User Interaction	Exploit	Availability Impact
Wesleyan University	Education	High	BAC net	CVE-2019-12480	Network	Low	None	None	EDB-ID-47148	High
Drexel University	Education	High	BAC net	CVE-2019-12480	Network	Low	None	None	EDB-ID-47148	High
Nova Scotia Dept. of Education	Education	High	BAC net	CVE-2019-12480	Network	Low	None	None	EDB-ID-47148	High
Helena Public Schools	Education	High	BAC net	CVE-2019-12480	Network	Low	None	None	EDB-ID-47148	High
Chery Creek School District	Education	High	BAC net	CVE-2019-12480	Network	Low	None	None	EDB-ID-47148	High

The 155 results included various types of education organizations across the world such as:

- K-12 school districts
- Public and private K-12 schools
- Universities and colleges
- Higher education networks
- University consortiums
- Education departments

The 155 distinct education organizations from the complex SPARQL query represent education organizations that have highly exploitable ICS assets that could be attacked in such a way that the ICS assets can no longer function. Attackers can shut down systems such as physical security systems, heating systems, ventilation systems, fire suppression systems, and other types of ICS systems in K-12 schools, universities, and colleges. Attacks on these ICS systems can significantly disrupt education organizations which can cause a “debilitating effect” overall (CISA, 2020).

In our research project, we started off with 126,636 ICS targets (a Shodan dataset of IP addresses that use ICS protocols) and then used NLP to reduce that list to 2,272 education organizations. However, a list of 2,272 potential targets can still result in several hours of work for ethical hackers in finding highly exploitable education targets. With a complex SPARQL query, we reduced the original target list even further to 155 distinct education organizations that have highly exploitable vulnerabilities with known exploits that can lead to the denial-of-service to critical ICS devices. This reduced the initial Shodan dataset of 126,626 targets by 124,354 targets (98.2%).

Overall, our research showed how semantic-web technologies (ontologies, NLP, and semantic queries) can be used to complete a situation assessment task in ethical hacking (evaluating a list of targets to find education organizations that have highly exploitable vulnerabilities with known exploits). We showed how semantic-web technologies can reduce the information overload faced by cybersecurity professionals (D'Amico et al., 2005; Tyworth et al., 2012; Yen et al., 2010). We also showed how semantic-web technologies can synthesize information together to solve the puzzle of finding targets with specific attributes, all of which represent a situation assessment task (Ben-Bassat and Freedy, 1982). Our results build on previous research that shows how cybersecurity professionals can fuse data from multiple sources to make informed decisions in cyber security (Barford et al., 2010; D'Amico et al., 2005).

## **5. Conclusion and future work**

Evaluating high-value targets based on organization-type, asset-type, and vulnerability data represents a complex situation assessment task for ethical hackers. Initially, we found over 126,000 potential targets that use well-known ICS protocols. With ontology modeling, NLP, and SPARQL queries, we focused on developing reasoning to find education-targets (an example of critical infrastructure targets) that use ICS protocols which are highly susceptible to severe cyber-attacks and that have publicly known exploits. We narrowed down the target list to 155 distinct education targets based on that reasoning. Although our research specifically focused on education targets, the techniques that we utilized in our research can be applied to other critical infrastructure domains.

In future research, we will examine how NLP can be used to identify other types of critical infrastructure targets (such as government organizations) and to evaluate the target value of other types of assets besides ICS assets. To evaluate target values of assets, we will examine how to incorporate various cyber risk frameworks such as the National Institute of Standards and Technology (NIST) Risk Management Framework (RMF). Evaluating various types of assets will be critical as ethical hackers could find thousands of assets when scanning a single organization. Overall, our research shows how ontologies can be used to automate situation assessment tasks of ethical hackers as they search for high-value targets including critical infrastructure organizations that use highly exploitable ICS assets.

## **Acknowledgements**

This research is supported, in part, by National Science Foundation Grant No. 1922202, CyberCorps Scholarship for Service (SFS).

## **References**

- Angles, R. and Gutierrez, C., 2008, October. "The Expressive Power of SPARQL". International Semantic-web Conference.
- Aviad, A., Wecl, K. and Abramowicz, W., 2015, July. "The Semantic Approach to Cyber Security towards Ontology Based Body of Knowledge." European Conference on Cyber Warfare and Security.
- Balducci, M., Kushner, S. and Speck, J., 2015, June. "Ontology-driven Data Semantics Discovery for Cyber-security." International Symposium on Practical Aspects of Declarative Languages.
- Barford, P., Dacier, M., Dietterich, T.G., Fredrikson, M., Giffin, J., Jajodia, S., Jha, S., Li, J., Liu, P., Ning, P. and Ou, X., 2010. "Cyber SA: Situational Awareness for Cyber Defense." In *Cyber Situational Awareness* (pp. 3-13). Springer, Boston, MA.
- Ben-Bassat, M. and Freedy, A., 1982. "Knowledge Requirements and Management in Expert Decision Support Systems for (Military) Situation Assessment." *IEEE Transactions on Systems, Man, and Cybernetics*, 12(4), pp.479-490. CISA, 2020. "Critical Infrastructure Sectors." [online] [www.cisa.gov/critical-infrastructure-sectors](http://www.cisa.gov/critical-infrastructure-sectors).
- D'Amico, A., Whitley, K., Tesone, D., O'Brien, B. and Roth, E., 2005. "Achieving Cyber Defense Situational Awareness: A Cognitive Task Analysis of Information Assurance Analysts. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 3, pp. 229-233).
- Elliott, T., 2005. "Expert Decision-making in Naturalistic Environments: A Summary of Research."
- Endsley, M.R. and Garland, D.J. eds., 2000. *Situation Awareness Analysis and Measurement*. CRC Press.
- Falk, C., 2016. "An Ontology for Threat Intelligence." European Conference on Cyber Warfare and Security.
- FBI, CISA, and MS-ISAC, 2020. "Cyber Actors Target K-12 Distance Learning Education to Cause Disruptions and Steal Data." [online] [us-cert.cisa.gov/ncas/alerts/aa20-345a](https://us-cert.cisa.gov/ncas/alerts/aa20-345a)
- Flowers, A.S., Smith, S.C. and Oltramari, A., 2016. "Security Taxonomies of Industrial Control Systems." In *Cyber-security of SCADA and Other Industrial Control Systems* (pp. 111-132). Springer, Cham.
- Freund, J. and Jones, J., 2014. *Measuring and Managing Information Risk: A FAIR approach*. Butterworth-Heinemann.
- Goodall, J.R., Lutters, W.G. and Komlodi, A., 2009. "Supporting Intrusion Detection Work Practice." *Journal of Information System Security*, 5(2), pp.42-73.

- Gorodetsky, V., Karsaev, O. and Samoilov, V., 2005. "On-line Update of Situation Assessment: A Generic Approach." *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 9(4), pp.351-365.
- Grant, T., van't Wout, C. and van Niekerk, B., 2020. "An Ontology for Cyber ISTAR in Offensive Cyber Operations" *European Conference on Cyber Warfare and Security*.
- Gruber, T.R., 1993. "A Translation Approach to Portable Ontology Specifications." *Knowledge Acquisition*, 5(2), pp.199-220.
- Kirillov, V.P., 1994. "Constructive Stochastic Temporal Reasoning in Situation Assessment". *IEEE transactions on Systems, Man, and Cybernetics*, 24(8), pp.1099-1113.
- Lipshitz, R. and Ben Shaul, O., 1997. "Schemata and Mental Models in Recognition-primed Decision Making." *Naturalistic Decision Making*, pp.293-303.
- Morris, T.H. and Gao, W., 2013, September. "Industrial Control System Cyber-attacks." *1st International Symposium for ICS & SCADA Cyber Security Research 2013 (ICS-CSR 2013)* 1 (pp. 22-29).
- Noble, D., 1993. "A Model to Support Development of Situation Assessment Aids." *Decision Making in Action: Models and Methods*, pp.287-305.
- Noble, D.F., Boehm-Davis, D. and Grosz, C., 1986. "Schema-based Model of Information Processing for Situation Assessment." *Engineering Research Associates, Inc., Vienna, VA*.
- Pérez, J., Arenas, M. and Gutierrez, C., 2006, November. "Semantics and Complexity of SPARQL." *International Semantic-web Conference* (pp. 30-43).
- Rogers, G. and Ashford, T., 2015. "Mitigating Higher Ed Cyber-attacks." *Association Supporting Computer Users in Education*.
- Salas, E., Rosen, M.A. and DiazGranados, D., 2010. "Expertise-based Intuition and Decision Making in Organizations." *Journal of Management*, 36(4), pp.941-973.
- Serfaty, D., MacMillan, J., Entin, E.E. and Entin, E.B., 1997. "The Decision-making Expertise of Battle Commanders." *Naturalistic Decision Making*, pp.233-246.
- Takahashi, T. and Kadobayashi, Y., 2015. "Reference Ontology for Cybersecurity Operational Information." *The Computer Journal*, 58(10), pp.2297-2312.
- Tyworth, M., Giacobe, N. A., Mancuso, V., and Dancy, C. (2012). "The Distributed Nature of Cyber Situation Awareness." In *IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*.
- Yen, J., McNeese, M., Mullen, T., Hall, D., Fan, X., and Liu, P. (2010). "RPD-based Hypothesis Reasoning for Cyber Situation Awareness." In S. Jajodia (Ed.), *Cyber Situational Awareness* (pp. 39–49). Springer.
- World Wide Web Consortium, 2013. "SPARQL 1.1 Overview." [online] <https://www.w3.org/TR/sparql11-overview/>
- World Wide Web Consortium, 2015. "Semantic Web." [online] from <https://www.w3.org/standards/semanticweb/>

# Educating the Examiner: Digital Forensics in an IoT and Embedded Environment

Iain Sutherland<sup>1</sup>, Huw Read<sup>1,2</sup> and Konstantinos Xynos<sup>1,3</sup>

<sup>1</sup>Noroff University College, Agder, Norway

<sup>2</sup>Norwich University, Northfield, USA

<sup>3</sup>MycenX Consultancy Services, Stuttgart, Germany

[iain.sutherland@noroff.no](mailto:iain.sutherland@noroff.no)

[hread@norwich.edu](mailto:hread@norwich.edu)

[kxynos@mycenx.com](mailto:kxynos@mycenx.com)

DOI: 10.34190/EWS.21.041

**Abstract:** The Internet of Things (IoT) is an interconnected world of semi-autonomous systems capable of automation, communication and monitoring. It encompasses all manner of systems and embedded devices, communicating using various protocols and standards. Sometimes these devices are purpose built for commercial or industrial environments and at other times generic builds provide domestic solutions. These systems have the potential to hold a significant amount of information on user preferences and activities as well as on the surrounding environment. Some data will usually reside on the device itself, or as seen in many cases, within a manufacturer supported cloud solution. Mobile and web applications will then provide a way to interface with the data or the device. The question arises as to the readiness of the Digital Forensic Examiner. There is a requirement to correctly identify the value of IoT and embedded systems at the crime scene. Once identified, the examiner needs the skill and knowledge to access, interpret and present the information that may be contained in this ever-expanding wide variety of devices found in home and work environments. The skills required by the Digital Forensic Examiner has progressed further from the analysis of hard drives and file systems. It must address the growing demand and requirement to be able to understand how embedded firmware operates. In some cases, this includes interpreting embedded code, memory structures and proprietary file formats. This paper discusses the increasing complexity of the changing environment. It reviews the types of skills and training needs and the subject areas for consideration when training forensics examiners over the next five to ten years.

**Keywords:** digital forensics, IoT, investigative strategy, training, skill sets

---

## 1. Introduction

The increase in the use of smart technology or smart devices (composed of an embedded processor, memory and a communication channel) is expanding the interaction between users and the variety of digital devices. There has been a migration away from user data and user activities being held just in laptop and desktop systems. A user may now interact with a range of devices with capabilities that may vary from simple environmental monitoring to more complex information gathering and processing. These environments have the potential to capture a user's activities in increasing detail, potentially providing information on timing, location and preferences. This information will be of interest to an examiner in determining a user's actions or whereabouts. These devices are frequently designed to carry out a specific, often limited activity but the user may be unaware of the amount of data gathered by the device. Examiners have an interest in access to any potential sources of information that may assist them in their casework. In terms of abilities, an investigative team would need to know many different systems/environments and have prior access in order to know what devices to examine and which sources will provide the greatest opportunity for intelligent, actionable information.

## 2. The environment

Evidence to date (Sarris et al, 2020) regarding the development of complex software systems suggests that integrating numerous smaller components creates concern with regard to updates and ongoing maintenance. The complexity of software leads to bugs and security flaws, while the interaction between software can lead to unintended functionality; for example MITRE's CWE (Mitre, 2020) list focuses on identifying weaknesses that lead to vulnerabilities that may be exploited by cyber criminals. The ever-present complexity of software is an indication of the potential problems to come with more complex Internet of Things (IoT) environments. Security flaws in these devices can also be exploited by malware such as Mirai (Zhang et al, 2020) and Reaper (TrendMicro, 2017) spreading specifically through IoT systems.



Examples of these emerging environments are the continuously evolving smart homes which are causing concern on privacy (Bronshiteyn, 2020), in particular with examples like the always on and listening Amazon Alexa (Shackleton, 2019). The need to analyse the behaviour of such systems is becoming increasingly apparent. As noted in the Interpol Review of Digital Evidence 2016-2019 (Reedy, 2020), criminals are early adopters of technology, an example being Smart TV's (Sutherland et al, 2014) and the subject of a search warrant in 2015 (Brewster, 2017, Bronshiteyn, 2020). Cyber criminals may also be unwittingly capturing information on illegal / illicit activity, allowing examiners to take advantage of the systems, as noted by Bronshiteyn (2020) and Shackleton (2019).

### **3. The challenges and needs of the digital forensic examiner**

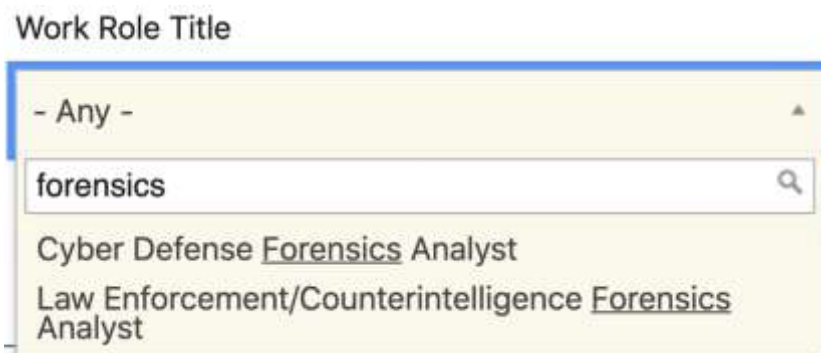
The digital forensic examiner is no longer solely concerned with laptop and desktop systems and the 'traditional' Windows, MacOS and Linux operating systems and associated file systems commonly used on these devices. The popularity of mobile phones introduced mobile-oriented OS, such as iOS and Android, as other important operating systems and possible sources of information. IoT devices typically include some form of control device, often a smartphone used to communicate with the device or a central hub. Although smartphones can provide some information either stored locally and presented in the control application, or from data stored in a cloud storage system, it has been shown that these different views of data (on-device, cloud, app) can result in varying quality of evidence (Awasthi, 2018). The challenge is that relying on the application may not provide access to all of the useful data on the device. While many IoT devices, like other embedded systems, often run versions of the Linux and more recently Windows 10 IoT operating systems, the actual IoT devices tend to have very limited connectivity and may also run proprietary code with no permanent file system. Accessing and interpreting the data will require a very specific approach (Watson, 2016) for which traditional guidance (e.g. ACPO see Horsman, 2020) is unsuitable.

There are obvious risks to limiting analysis to just the control application on the smartphone and the point and click options of the digital forensic tool set. Investigating this type of environment increasingly requires an in-depth understanding of devices and potentially components at a hardware level.

### **4. Towards defining a skill set**

#### **4.1 Guidelines, standards and skill sets**

Discussions addressing digital investigation challenges tend to have a government and industry focus, with a limited emphasis on the role of academia. Indeed, in recent work by Casey et al, (2019) which determine 4 goals for mitigating such challenges, academia is not mentioned. In the United States of America, the NICE (NICCS, 2020) workforce framework has been publicly available since 2017, with its most recent revision in NIST (2020). It is moving towards describing the broad spectrum of cybersecurity work; one of the goals is to allow educational establishments and employers to describe the roles in a common language. "Work roles" are made of "tasks" which require "knowledge" ("a retrievable set of concepts within memory") and "skills" ("the capacity to perform an observable action"). An employee's ability to do the job, or a student's ability to demonstrate proficiency, can be assessed via "competencies" which test the person's "knowledge" and "skills", thereby their ability to perform "tasks" for the "work role". Of particular interest are the predefined "work roles" relating to digital forensics (Fig. 1, NICCS, undated).



**Figure 1:** NICE workforce framework work roles for digital forensics

For both of these digital forensic focused work roles (Fig. 1.), a number of skills and knowledge items are presented, along with a number of capability indicators for entry, intermediate and advanced levels. Whilst there is no mention of embedded or IoT systems specifically, and it is unclear how often these work roles are updated to reflect changes in technology (no timestamps on the webpage), the framework does have the flexibility to incorporate additional tasks as identified by an employer or an educator. However, it should be highlighted that whilst the NICE framework does not specifically provide guidance for digital forensic academic programmes, it is possible to infer what may be expected of one by examining the work tasks.

In the UK a programme was developed for certifying masters degrees in digital forensics. GCHQ defined in 2015 an outline set of criteria that a digital forensics degree needed to meet in order to achieve their certification. This programme is now maintained by the NSCC (2019) as an Integrated Masters with a specific Digital Forensics Pathway. This document has a series of appendices that refers to the specific subject areas of both the core computer science sections and the specific subject pathways (including digital forensics). The indicative topics are described at a very high level. The core computer science programme requires coverage of embedded systems; debugging including JTAG and UART and side channel analysis. It also requires low level techniques and tools for malware analysis and reverse engineering. There is a mention of mobile devices in the forensics element of the degree programme. This document is dated in early 2019 and highlights the increasing importance of embedded systems although there is no specific mention of IoT devices.

The international standard on ISO/IEC 27037:2012 “Information technology — Security techniques — Guidelines for identification, collection, acquisition and preservation of digital evidence”, defines some key skills for the Digital Evidence first responder. In terms of recognition this is focussed on networking concepts, while the acquisition of evidence is broader and mentions “RAID, database, appliances and miniaturized devices;” (ISO/IEC 27037:2012). A related standard, ISO/IEC 27042:2015 “Information technology — Security techniques — Guidelines for the analysis and interpretation of digital evidence”, takes a different approach in defining how competencies and proficiencies can be described. The international standards date back to 2012 and 2015 and as standards, as to be expected, are focussed on the practice and procedure rather than defining the skills required to complete the investigative tasks.

#### **4.2 Digital forensic laboratories and skill requirements**

One further area to explore is the documentation and standards for creating digital forensics laboratories which, in addition to outlining laboratory requirements, also comment on the basic skill sets or roles for digital forensic examiners. There are a number of documents looking at forensic best practice (ACPO 2012a), laboratory operations and staffing of digital forensic laboratories including ACPO (2012b). These could provide some basis for determining the required skill set from the perspective of operating a laboratory. The ACPO lab managers guide ACPO (2012b) defines the following roles and responsibilities within a computer crime unit:

- 1. Administration / Reception Officer
- 2. Triage Officer
- 3. Physical examiner/imager
- 4. Previewing Examiner
- 5. Advanced Examiner
- 6. Quality Assurance Officer
- 7. Training Officer

ACPO roles of interest are: Triage Officer (filtering cases), the Physical examiner/imager (identification imaging), the Previewing Examiner (automated processes and bespoke) and the Advanced Examiner (Specialist examination). These would all require specific knowledge in their areas. The best practice guidelines have the same issue as the international standards in that they date back to 2012. Considering the pace of change in the forensics environment, both of the ACPO guides (2012a, 2012b) in forensics terms are somewhat dated, and there have been suggestions that the principles and practices at least need revision (Horsman, 2020). Reedy (2020) also notes that “...best practice guidelines were established over a decade ago and do not meet the challenges of smart technology, and some do not address memory forensics, database forensics, or network forensics...”. A more recent guide defining the development of a forensics laboratory (INTERPOL, 2019) outlines the requirements for an examiner as shown in Figure 2, highlighting the laboratory specific nature of an

examiner. An examiner would inevitably develop the skill set for the tasks commonly covered in the laboratory in which they are employed.

**Examiner**  
The Examiner must have the relevant technical knowledge and appropriate qualifications. Ideally he/she should have some training in the use of DF software.

On being hired, the Examiner will be required to attend specific training to obtain a minimum set of skills. The Examiner must have knowledge of legislation and be aware of the elements of each offence in order to articulate those facts when investigating different types of crimes. These roles require an analytical and investigative mind-set. The Examiner must also be able to deliver his/her findings in a clear and understandable manner, therefore having good oral and written communication skills is essential.

**Figure 2:** INTERPOL (2019) description of an examiner role

The document (INTERPOL, 2019) also outlines a skill set list for the Digital Forensics Examiner; data recovery, computer forensics, mobile phone forensics, audio video and image forensics and emerging technologies. Considering the DFL skill set Checklist (Appendix A of INTERPOL, 2019), a number of skills are identified, the closest related to IoT being the *emerging technology* category. However, such embedded systems are not a standalone category, as the knowledge and skills required for embedded systems would also be needed to extract evidence from devices in some of the other categories listed in the framework. The list of emerging technologies includes vehicle, shipborne, drone and other 'new' technologies. The skill set required is the need to access, extract and interpret data from a range of embedded computing devices, clearly a need to cover more than traditional computing technologies. The document also defines aspects of the proposed roles including a technical understanding of data storage, microchip operation and basic electronics in addition to practical skills and possibly most critical logical thinking and analytical skills (INTERPOL, 2019). JTAG, Chip-off, rooting and Jailbreaking are all skills now required by the digital forensic examiner (INTERPOL 2019). Therefore, an increasing broad skill set, and a greater depth of knowledge is needed by the forensic examiner, unless there is a degree of specialization separating digital forensics into specific knowledge, skills and abilities to deal with IoT, triage in the crime scene and other niche requirements (i.e. cloud-based acquisition etc).

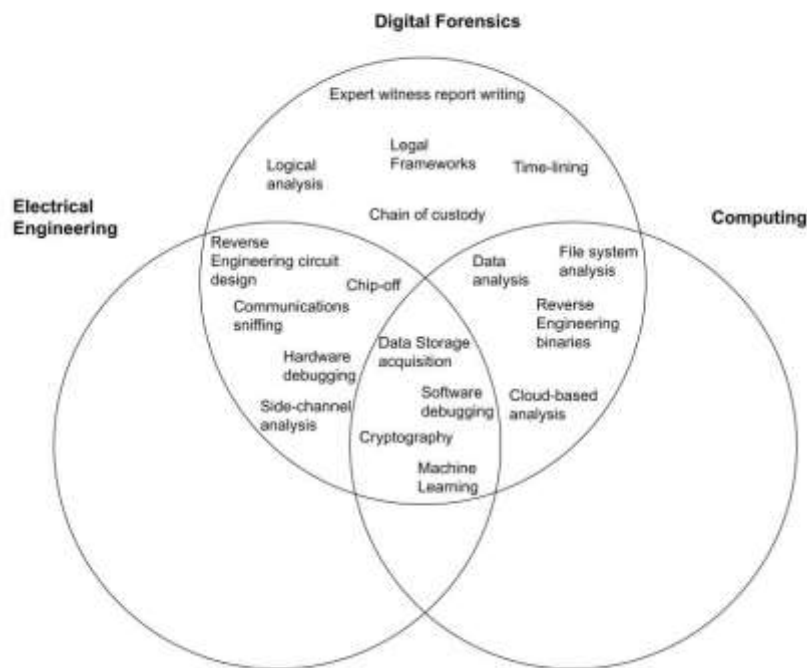
## **5. Training the digital forensics examiner**

Digital forensic examiners in post are recommended to undergo training and, or certification in the tools they are using (INTERPOL, 2019). There are a number of university undergraduate programmes in place to educate new digital forensic examiners. These programmes face a number of challenges. Reedy (2020) notes that "*digital forensics is becoming inaccessible due to the increasing expense and complexity...*". Cost is clearly also a factor for universities when considering new study programmes. A number of papers exploring curriculum development including Tu et al (2012) and notably Lang et al (2014), outline several key areas for concern in developing a forensics curriculum. The following sections highlight a selection of the challenges and concerns raised in Lang et al (2014) and reflect on IoT and embedded challenges.

### **5.1 Digital forensics curriculum challenges**

One of the most significant issues and possibly greatest opportunities can be argued to be the lack of a specified digital forensics curriculum design. There are a number of perspectives of what should be included in a digital forensics curriculum. NIST have proposed a framework for the development of cyber security materials via NICE (NIST, 2020). In addition to Tu et al (2012) and notably Lang et al (2014), there has been other work in the area including Srinivasan (2013) exploring integrating digital forensics into a security curriculum. The general consensus is the applied nature of the program, but also the need to learn both theory and skills in a number of topics, as well as having test devices at hand to practice the processes and procedures.

The addition of an IoT skill set poses a challenge (Yaqoob et al, 2019) in further extending what may already be a full curriculum. For example, the need to understand how to extract contents directly from solid state storage on a PCB and interpret the contents is highly-specialised, and requires several electrical engineering, computing and digital forensic-specific prerequisites in order to understand the implications of such low-level analysis. Fig. 3 provides a focussed insight into the overlaps that occur between the three domains of forensics, electrical engineering and computing. Therefore, as a simplification the problem can be split into two realms, hardware and software.



**Figure 3:** Relationship of selected key skills (focusing on digital forensics)

When dealing with the hardware there is a need for some knowledge in electrical engineering. This would cover some basic concepts required when dealing with hardware related challenges of a case. This requirement is not limited to IoT devices, since it is seen already in mobile device forensics (e.g., chip-off) and CCTV systems. The main advantage of having physical access and interrogating IoT devices, and other such related devices, in such a way is that it provides more opportunities for interacting with the device and the OS. Interactions could involve debugging interfaces (e.g. kernel boot argument manipulation, JTAG, UART, terminal/SSH), sniffing of communication channels (e.g. SPI, CAN, RFID, Wifi, BLE, Zigbee etc), networking capabilities (e.g. Wifi, BLE, Zigbee etc) and even accessing storage mediums (e.g. memory cards, storage chips (SPI flash, NAND, eMMC), EEPROM etc.). Reverse engineering circuit design is just as important in identifying device capabilities and data sources.

Once the hardware issues have been addressed and the device's software and data has been dumped, the focus shifts to the software. The analyst will need to know how to overcome the formats of the extracted data (e.g., related or specific to the hardware device and medium involved) in order to proceed. From then on the problem is one very familiar to the digital forensic analyst. In most cases the device's file system will be extracted along with configuration details, and where present, saved data. An understanding of the type of OS used by embedded systems (e.g. Linux, Android, QNX etc.), processor architectures (e.g. MIPS, ARM), possible cloud-based APIs, mobile control devices and some binary reverse engineering would also be needed. This knowledge enables the investigator to understand sources of data and possible changes that would occur on the system being investigated. This would create specialists in IoT digital forensics that are able to understand and approach the system depending on the individual requirements of an investigation.

Combining hardware and software knowledge can be seen as follows: If an analyst wanted to boot a system as root and bypass the login prompt they could implement the following; they could dump the system using a hardware device (e.g. SPI), make a modification to the kernel boot arguments that force the system to boot as root and write the flash back to the system. This would involve knowledge of hardware and software capabilities of a device. The forensic value of this type of approach is open to debate since the device would have to be shut down, modified and started again, therefore losing any volatile data. With root access to the system, it would then be possible to extract forensic images of the live system as and when needed, including volatile artifacts.

## **5.2 Teaching and delivery resources**

There are a variety of textbooks available for teaching digital forensics. However, the process of creating textbooks can take a number of months and could lag behind as the subject rapidly evolves. Most educational institutions would combine core textbooks with research papers to provide the latest and most up to date material. Textbooks are also written for the largest markets, typically the USA and to a lesser extent Europe. This is a challenge as digital forensics is a combination of computing and law, so textbooks with legal system information are often inappropriate due to jurisdictional differences. Staff training and continued development in the aforementioned areas is important, although not without difficulty. This is especially true when a computing graduate would have to be aware of some electrical engineering topics. Recruiting staff with this specialised knowledge and experience would be even more of a challenge.

Given the disparate nature of devices considered to be part of the digital forensics discipline, existing corpora such as Digital Corpora (2021), cannot be expected to provide samples of all devices. Teaching images are time consuming to develop, and so need to be reused due to the degree of effort involved in creating the material. However, studies have shown (Carthy et al, 2018 and Woods et al, 2011) it is possible to integrate the development of such material into the pedagogical outcomes of digital forensic courses, but these have focused on familiar technology/media rather than exploring possibilities with new devices. Focused challenges help with the direction of digital forensic research (e.g. DFRWS, 2018) but less so with the educational aspects of the discipline.

While data recovery may be addressed in some forensics degrees, the inclusion of data extraction and IoT / embedded systems introduces another dimension in the generation of appropriate case studies and practical sessions.

## **5.3 Jurisdictional issues**

There are several challenges when considering legal aspects and how they relate to digital forensics. Local laws and customs relative to where a student is studying, geographic location of crime, location of digital evidence and location of organisation holding data (e.g., CLOUD Act, 2018) all make it challenging to provide an appropriate in-depth review of legal issues.

### *5.3.1 Laboratory resources*

Core digital forensics concepts can be taught with standard computing facilities and open-source software tools. However, to cover all of the stages of an investigation and to provide the required practical skills and experience, some considerable investment is required. The initial investment cost may be prohibitive to most academic institutions when factoring in the specialist hardware, software licenses, staff training and continuing professional education, not to mention the required student computers. The laboratory may share some similarities with a cyber security laboratory, the benefit of a closed network with local storage for shared resources. If the laboratory is running commercial software locally, these computers are likely to be more expensive than those in a standard university laboratory.

If a laboratory is expected to examine IoT devices and other embedded systems, then there are additional expenses. Hot air rework stations, air handling systems, dismantling benches and associated tools are not uncommon. Specialist analysis hardware such as Saleae Logic Analyzer, Hardsploit, JTAGulator, BusPirate and Hydrabus and associated consumables (cables, clips) in addition to staff training and the development of safety policies and procedures would all be required.

## **5.4 Changing environment for the delivery of courses**

The COVID-19 pandemic in 2020 and 2021 has seen many higher education institutions move to online delivery. This poses a challenge for any degree programme where practical elements are included, but especially with applied courses. Simulating a network and associated security problems in a series of virtual machines is possible. Gaining practical experience in extracting the contents of an SPI chip or mobile device using specialist hardware is not possible. The shift online increases the likelihood of enrolling international students, further exacerbating the challenges relating to legal frameworks and different jurisdictions.

## **5.5 Discussion and conclusions**

The increasing connection between cyberspace and the physical world resulting from IoT and embedded devices presents an increasingly complex, evolving environment for Digital Examiners. The volume of data that could potentially be useful is expanding, but so is the distribution of that data across increasingly varied devices and systems. The Examiner, in addition to 'traditional' multiple operating and file systems, also has to process embedded code and devices with more esoteric file systems. The increasing importance of network forensics and the need to deal with multiple devices communicating over a variety of different protocols means the Digital Examiner will need to conduct triage on-site, as well as be aware of the device's capabilities. While Hitchcock et al (2016) argued that triage could be undertaken on site by a non-specialist, there was also an acknowledgement that training would be required for the more advanced tasks such as memory capture. The interconnected nature of the systems and the volatility of data means that evidence dynamics will also be an increasing feature of the environment.

It is possible to derive a number of conclusions. The emerging technologies referred to in appendix 1 in INTERPOL, (2019) which in addition to social media, database, cryptocurrency and biometric technologies highlighted shipborne, vehicle and drone systems. The latter are likely to include embedded systems, but so will other INTERPOL categories, in particular audio and video systems (Martin et al, 2021). In exploring the current environment and the challenges facing the digital forensic examiner, and having reviewed some of the issues with creating a suitable digital forensics curriculum, it is clear that digital forensics training will now need to address a number of areas:

- 1. An understanding of the technologies in an embedded system, in particular the communication standards and protocols used which could be JTAG, SWD, UART, USB, I2C, SPI, etc that enable extraction of data directly from the physical device. These require an additional knowledge base, to extract information from the device.
- 2. An ability to use appropriate tools to examine bus communications and data stored in SPI flash, NAND and other PCB mounted components. In addition to write-blockers, other tools would be required, but not limited to; Saleae Logic Analyzer, Hardsplloit, JTAGulator, BusPirate and Hydrabus. These will become part of the digital forensics toolkit as will the necessary skills to use these tools to extract intelligible data from IoT devices and embedded systems. The investigator, during initial evidence seizure and triage, must be competent with such tools and methods to be prepared to capture volatile data on-site (Zulkipli, 2021).
- 3. An ability to interpret the findings from disparate devices and place them in comparative context within the environment. To have the competency to interpret the evidence dynamics of the system and the relative importance of the findings. In the longer term, incorporating advanced Machine Learning systems (i.e., Huybrechts et al, 2018) that help identify custom binaries and protocols, and support the new processes and procedures of working with IoT and embedded devices.
- 4. The continuing evolution of technology focussed on securing information in the device using features such as Physical Unclonable Functions (PUFs) (Gao, 2020) will only make accessing data more challenging. In addition, the increasing adoption of device encryption means Digital Forensics Examiners will also ideally need to collaborate closely with IoT and embedded device manufacturers, who in turn will have to implement forensic readiness methods as part of the physical device's life cycle.

The low-level understanding required to analyse firmware and to access systems via a variety of methods is essential. As technology moves to fully adopt cryptographic elements and enhanced device security (ENISA, 2017), it is highly possible that security specialists, such as cryptographers and a cryptoanalysis team, would also be part of the future team of Digital Forensics Examiners. When considering the broad range of skills needed for the Digital Forensics Examiner, it is clear that there are significant challenges when, for example, creating a university degree. A series of specialist forensics topics, building on a computer science degree may no longer be sufficient. The range of required skills has broadened to cover aspects of computing, electronics and law.

## **Acknowledgements**

Special thanks to Septimiu Mare and Andrew Blyth for some insightful conversations.

## **References**

ACPO, (2012a) Good Practice Guide for Digital Evidence, Police Central E Crime Unit, March 2012 [https://www.digital-detective.net/digital-forensics-documents/ACPO\\_Good\\_Practice\\_Guide\\_for\\_Digital\\_Evidence\\_v5.pdf](https://www.digital-detective.net/digital-forensics-documents/ACPO_Good_Practice_Guide_for_Digital_Evidence_v5.pdf)

- ACPO, (2012b) Good Practice and Advice Guide for Managers of e-Crime Investigation, Official Release Version V0.1.4 <http://www.acpo.police.uk/documents/crime/2011/201103CRIECI14.pdf> Now available at - [https://www.digital-detective.net/digital-forensics-documents/ACPO\\_Good\\_Practice\\_and\\_Advice\\_for\\_Manager\\_of\\_e-Crime-Investigation.pdf](https://www.digital-detective.net/digital-forensics-documents/ACPO_Good_Practice_and_Advice_for_Manager_of_e-Crime-Investigation.pdf)
- Awasthi, A., Read, H.O., Xynos, K. & Sutherland, I., (2018) Welcome pwn: Almond smart home hub forensics. *Digital Investigation*. 2018, 26, 38–46.
- Brewster T., (2017) That Time Cops Searched A Samsung Smart TV For Evidence Of Child Abuse. Available online at: <https://www.forbes.com/sites/thomasbrewster/2017/02/07/samsung-smart-tv-fed-search-child-pornography/?sh=5a25469417d7> Last Accessed 22 Feb 2021.
- Bronshsteyn G., (2020) Searching the Smart Home. *Stanford Law Review*, February 2020, Volume 72, Available online at: <https://review.law.stanford.edu/wp-content/uploads/sites/3/2020/02/Bronshsteyn-72-Stan.-L.-Rev.-455.pdf>
- Carthy L., Little R., Øvensen E., Sutherland I. & Read H.O.L., (2018) Committing the perfect crime: A teaching perspective. 17<sup>th</sup> European Conference on Cyber Warfare and Security 28 - 29 June 2018, Oslo, Norway.
- Casey, E., Geradts, Z. & Nikkel B., (2019) Editorial: Transdisciplinary strategies for digital investigation challenges, *Digit. Invest.* 25 (2019) 104.
- CLOUD Act., (2018) Clarifying the Lawful Use of Overseas Data Act of 2018, Pub. L. No. 115–141, 132 Stat. 348 (codified as amended in separate sections of 18 U.S.C.); available online at <https://cli.re/BwPk5Q>
- DFRWS, (2018) DFRWS 2018 challenge. Available online at: <https://github.com/dfrows/dfrows2018-challenge>
- Digital Corpora., (2021) Producing the digital body. Available online at: <https://digitalcorpora.org/>
- ENISA, (2017) Baseline Security Recommendations for IoT, Available online at: [https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot/at\\_download/fullReport](https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot/at_download/fullReport)
- Gao, Y., Al-Sarawi, S.F. & Abbott, D., (2020) Physical unclonable functions. *Nat Electron* 3, 81–91 <https://doi.org/10.1038/s41928-020-0372-5>
- Hitchcock B., Le-Khac N-A. & Scanlon M., (2016) Tiered forensic methodology model for Digital Field Triage by non-digital evidence specialists. DFRWS 2016 Europe Proceedings of the Third Annual DFRWS Europe, Digital Investigation Volume 16, Supplement, 29 March 2016, Pages S75-S85
- Horsman, G., (2020). ACPO principles for digital evidence: Time for an update? *Forensic Science International: Reports*, 2, December 2020, [100076]. <https://doi.org/10.1016/j.fsisr.2020.100076>
- Huybrechts T., Vanommeslaeghe Y., Blontrock D., Van Barel G. & Hellinckx P., (2018) Automatic Reverse Engineering of CAN Bus Data Using Machine Learning Techniques. In: Xhafa F., Caballé S., Barolli L. (eds) *Advances on P2P, Parallel, Grid, Cloud and Internet Computing*. 3PGCIC 2017. Lecture Notes on Data Engineering and Communications Technologies, vol 13. Springer, Cham. [https://doi.org/10.1007/978-3-319-69835-9\\_71](https://doi.org/10.1007/978-3-319-69835-9_71)
- INTERPOL, (2019) Global Guidelines for Digital Forensics Laboratories. INTERPOL Global Complex for Innovation Available online at: [https://www.interpol.int/content/download/13501/file/INTERPOL\\_DFL\\_GlobalGuidelinesDigitalForensicsLaboratory.pdf](https://www.interpol.int/content/download/13501/file/INTERPOL_DFL_GlobalGuidelinesDigitalForensicsLaboratory.pdf)
- ISO/IEC 27037:2012, (2012) Information technology — Security techniques — Guidelines for identification, collection, acquisition and preservation of digital evidence
- ISO/IEC 27042:2015, (2015) Information technology — Security techniques — Guidelines for the analysis and interpretation of digital evidence.
- Lang A., Masooda B., Campbell R. & DeStefano L., (2014) Developing a new digital forensics curriculum. *Digital Investigation* 11, (2014) S76-S84
- Martin E D., Sutherland I. & Kargaard K., (2021) *IoT Security and Forensics: A Case Study*. ECCWS 2021, 20th European Conference on Cyber Warfare and Security, held at the University of Chester, UK
- MITRE, (2020), Common Weakness Enumeration: A Community-Developed List of Software & Hardware Weakness Types. Available online at: <https://cwe.mitre.org>
- NICCS, (undated) National Initiative for Cybersecurity careers and studies, NICE Cybersecurity Workforce Framework Work Roles <https://niccs.cisa.gov/workforce-development/cyber-security-workforce-framework/workroles>
- NIST, (2020) Special Publication 800-181 Revision 1 Workforce Framework for Cybersecurity (NICE Framework) this publication is available free of charge from: <https://doi.org/10.6028/NIST.SP.800-181r1> Also: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-181r1.pdf>
- NSCC, (2019) National Cyber Security Centre (a part of GCHQ) Certified Master’s in Cyber Security. Certification Of Integrated master’s Degrees (Computer Science and Digital Forensics (Pathway C) [https://www.ncsc.gov.uk/files/Certification-IntMasters-Issue-4\\_0-Feb-2019.pdf](https://www.ncsc.gov.uk/files/Certification-IntMasters-Issue-4_0-Feb-2019.pdf)
- Reedy P., (2020), Interpol review of digital evidence 2016 - 2019, *Forensic Science International: Synergy*, Volume 2, 2020, Pages 489-520, ISSN 2589-871X, Available online at: <https://doi.org/10.1016/j.fsisyn.2020.01.015>.
- Srinivasan S., (2013) Digital Forensics Curriculum in Security Education, *Journal of Information Technology Education: Innovations In Practice*, Volume 12, 2013 <http://www.jite.informingscience.org/documents/Vol12/JITEv12IIPp147-157Srinivasan1232.pdf>
- Sarris, D., Xynos, K., Read, H. & Sutherland, I., (2020) "Toward a methodology for the classification of IoT devices", European Conference on Cyber Warfare and Security 2020 (ECCWS), Chester, UK
- Shackleton, J., R., (2019). Alexa, Amazon Assistant or Government Informant?, 27U. *Miami Bus. L. Rev.*301, Available at: <https://repository.law.miami.edu/umblr/vol27/iss2/6>

***Iain Sutherland, Huw Read and Konstantinos Xynos***

- Sutherland, I., Read, H., Xynos, K., (2014). Forensic analysis of smart TV: A current issue and call to arms. *Digital Investigation*. 11. Issue 3 Sept. 2014 10.1016/j.diin.2014.05.019.
- TrendMicro, (2017) Millions of Networks Compromised by New Reaper Botnet. Available online at: <https://www.trendmicro.com/vinfo/pl/security/news/cybercrime-and-digital-threats/millions-of-networks-compromised-by-new-reaper-botnet> Last Accessed 22 Feb 2021.
- Tu, M. Xu D. & Cronin K., (2012) "On the Development of Digital Forensics Curriculum." *Journal of Digital Forensics Security and Law* 7 (2012): 13-32. DOI:10.15394/jdfsl.2012.1126 Corpus ID: 9263719
- Watson, S. & Dehghantanha, A., (2016) Digital forensics: the missing piece of the Internet of Things promise, (Feature) *Computer Fraud & Security*, Volume 2016, Issue 6, 2016, Pages 5-8, [https://doi.org/10.1016/S1361-3723\(15\)30045-2](https://doi.org/10.1016/S1361-3723(15)30045-2)
- Woods, K., Lee, C., Garfinkel, S., Dittrich, D., Russell, A. & Kearton, K., (2011) *Creating Realistic Corpora for Forensic and Security Education*, (2011) ADFSL Conference on Digital Forensics, Security and Law, 2011.
- Yaqoob, I., Hashem, I.A., Ahmed, A., Kazmi, S.M., & Hong, C., (2019). Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges. *Future Generation Computer Systems*, 92, 265-275. Available online at: <https://doi.org/10.1016/j.future.2018.09.058>
- Zhang, X., Upton, O., Beebe, N.L., & Choo, K.K.R., (2020) IoT Botnet Forensics: A Comprehensive Digital Forensic Case Study on Mirai Botnet Servers, *Forensic Science International: Digital Investigation*, Volume 32, Supplement, 2020, 300926, ISSN 2666-2817, Available online at: <https://doi.org/10.1016/j.fsidi.2020.300926>.
- Zulkipli N., H., N., Willis G.B. (2021) An Exploratory Study on Readiness Framework in IoT Forensics, *Procedia Computer Science*, Volume 179, 2021, Pages 966-973



# Interdependence of Internal and External Security

Ilkka Tikanmäki<sup>1,1</sup> and Harri Ruoslahti<sup>1,2</sup>

<sup>1</sup>Security and Risk Management, Laurea University of Applied Sciences, Espoo, Finland

<sup>2</sup>Department of Warfare, National Defence University, Helsinki, Finland

[ilkka.tikanmaki@laurea.fi](mailto:ilkka.tikanmaki@laurea.fi)

[harri.ruoslahti@laurea.fi](mailto:harri.ruoslahti@laurea.fi)

DOI: 10.34190/EWS.21.112

**Abstract:** Changes in the security environment, affecting both internal and external security, have been rapid in recent times. Security challenges related to hybrid phenomena, cybersecurity and organized cross-border crime significantly influence the development of the security environment. Global interdependence contributes to the nature of security, e.g., within the EU the free movement of goods and people have increased interdependence. The importance of situational awareness created and shared jointly by security actors is based on up-to-date information and assessments. Seamless cross-administrative collaboration promotes situational awareness (SA) and real-time situation picture. Thus, situational awareness is important for decision-making at different levels in various operating environments. Preparing for threats in accordance with the principle of total security is to safeguard the vital functions of society through cooperation between authorities, business, organizations, and citizens. Preparedness is a matter of comprehensive security and the vital functions in society involve cooperation between authorities, organizations, and citizens. As the operational environment is constantly changing, it has become increasingly difficult to distinguish between internal and external security and responding to changing threats may require revisions in policies and practices, and improved cooperation between actors. Significant changes in security situations may require addressing jurisdiction for security authorities and other actors, as jurisdiction is always based on the law. Effective cooperation between authorities requires responsible management, confidentiality, and appropriate allocation of resources. On an individual level, commitment, cooperative spirit, and personal contacts become critical to the success of collaborative work. The Common Operational Picture (COP) is a tool for achieving a good level of situational awareness, which in turn requires improved decision-making abilities and precise responses to situations that may arise. Positive developments are taking place in the field of information systems and information exchange between authorities. As threats change, so should the policies of states' internal and external security authorities be considered, also requiring reviewing the competences of these authorities, and how national legislation enables the security authorities to act in the face of possible threats.

**Keywords:** comprehensive security, internal security, external security, cooperation, situation picture, situation awareness, hybrid, common information systems

---

## 1. Introduction

Changes in the security environment have been rapid in recent times, affecting both internal and external security (European Commission, 2020). Security challenges related to organized cross-border crime, hybrid phenomena and cybersecurity will significantly influence the development of the security environment. The importance of situational awareness created and shared jointly by security actors is based on up-to-date information and assessments (Endsley, 2015). Threats and disturbances must be anticipated, prepared, and responded to.

Traditional security thinking has seen security as a separate entity that is only considered when a threat has already occurred. Changes in security situations are more likely in the future and it is difficult to prepare for them. (Ministry of Interior, 2017). In Finland, preparedness is looked as comprehensive security, the vital functions of society involves cooperation between authorities, organizations, and citizens. The relationship between external and internal security are closely interlinked (Prime Minister's Office, 2009; Hyvönen and Juntunen, 2021).

Terrorism and radicalization, cybercrime and illegal immigration pose challenges to the maintenance of internal security, while external security is affected by issues such as social and economic crises that do not respect national borders (Ministry of Interior, 2016). The Police Board has identified as a priority in internal security strategy the fight against organized crime, cybercrime and terrorism, and the fight against illegal immigration; as a means, the Police Board presented e.g. better exchange of information and cooperation with third countries (Police University College, 2020). "Preparedness and response for security threats require a strong national and

---

<sup>1</sup> <https://orcid.org/0000-0001-8950-5221>

<sup>2</sup> <https://orcid.org/0000-0001-9726-7956>

international co-operation, pre-agreed arrangements for cooperation between the authorities, business and NGOs." (Tuohimaa, Tikanmäki and Rajamäki, 2011, p. 611).

The research problem of this paper is to examine interdependencies between internal and external security, by looking at some of the main phenomena, threats, risks related to security, how are they influencing the security and preparedness and how has Finland considered them?

## **2. Hybrid threats and hybrid influence**

The Finnish Security Strategy for Society (YTS2017) defines hybrid engagement as an activity that "pursues its own goals through a variety of complementary means and by exploiting the weaknesses of the target". Means of hybrid influence can be economic, political, or military, and can be used simultaneously or sequentially with technology and social media. (The Security Committee, 2017.)

Hybrid influence can be divided into geo-economics, information, and electoral impact. Energy policy can be used as a tool for foreign influence (geo-economics), with cross-border energy transmission and imports. Trolls and cyber weapons can be used for information and electoral impact, based on a supranational IT infrastructure. (The Finnish Institute of International Affairs, 2018.)

Hybrid threats are not a new phenomenon for public authorities in Finland, the first Strategy for securing the vital functions of society discussed which authorities, business and organizations designed, prepared, and practised long-term responses to a wide range of security threats. (Finnish Government, 2003). Securing vital functions of society and managing overall security include preparing for threats, managing, and recovering from the disruptions and emergencies. Critical functions in society include leadership, international and EU action, defence capabilities, internal security, economy, infrastructure and security of supply, population capabilities and services, and mental resilience (Finnish Terminology Centre, 2017).

Hybrid threats can target vital functions and critical targets in society and may involve pressure, information operations and cyber operations (Järvenpää 2017). Due to varied and changing threats, authority policies in internal and external security need to be adapted to the new situation. Combating hybrid threats requires "an understanding of government, authorities, and industry to protect functions critical to decision-making and overall security in society. The best way to achieve this is through a coherent situation picture, the development of policies and practices, and training." (Lalu and Puistola, 2015, p. 4.) Resource sharing and resources common use are emphasized in preparing for hybrid threats (Uusipaavalniemi and Puistola, 2016).

General safety analysis plays a key role in identifying early forms of hybrid influence. As a result, the skills required by operators to detect and identify hybrid effects will increase. Actors should develop the necessary "capabilities and cyber security in cooperation with national and international actors". (Ministry of Interior, 2016, p. 13.)

A key objective of hybrid influence is to narrow the national sovereignty of another state. Mäkelä (2018, p. 13) describes hybrid influence as "a systematic activity in which a state or non-state actor can simultaneously use various military means or, for example, economic or technological pressure, as well as information operations and social media". The goal is to keep the hybrid effect at a level where it does not escalate into open conflict (Mäkelä, 2018.) Combating hybrid influencing requires the identification of one's own weaknesses, proactive preparation, situation awareness and situation understanding, clear procedures and leadership (Puistola, 2018). One defining characteristic is the continuous utilization of identifiable asymmetries, whether in the actual war or non-violent phase. Asymmetries are utilized as a combination of surprise, abuse, and deception. (Cederberg and Eronen, 2015.)

## **3. Interdependence of Internal and external security**

The boundary between external and internal, national, and international threats has become less clear, which affects the activities of security authorities (Prime Minister's Office, 2009). Threats can be divided into civilian and military in nature (McNeese et al., 2006), while civilian crisis management and military operations have come closer to one another, requiring civilian and military actors in both (Bendiek, 2017). The Ministry of Interior of Finland has identified that the main national threats and risks are large-scale uncontrolled influx of refugees, influencing energy networks and production, organised crime, and social exclusion, where the extreme

consequence of social exclusion can be radicalization and intensification of extremism (Ministry of Interior, 2017).

Global interdependence contributes to a further tightening of security, EU membership with its consequent free movement of goods and people, for example, have positively increased interdependence. The global sustainability crisis affects security, as it affects economy and well-being. There is a focus on climate change and the sufficiency of natural resources. The use of information networks is restricted by security and political considerations. Automation, artificial intelligence and robotization blur the interface between technology and humans (Ministry of Interior, 2017).

Disruptions to normal conditions can be dealt with existing jurisdictions of authorities, while significant changes in the security situation may require additional jurisdiction for security authorities and other actors. Jurisdiction is based on law, and while some jurisdictions are always valid, others can only be used under law crisis and during specifically defined situations. E.g. the jurisdiction of the Armed Forces is based on the Armed Forces Act, the police on the Police Act, and the Border Guard has special crime prevention functions, which are provided for in the Law on Crime Prevention at the Border Guard. (Finlex, 2020a, 2020b, 2020c.)

In Finland, cooperation between the police and judicial system has been natural and effective, and except for some coercive measures, the police have decision-making powers during operations. There is also “need to develop national legislation to match the current operational environment (Ministry of Interior, 2016, p. 73). Similar cooperation between the police and judiciary system is not possible everywhere, as national legislation in some countries limits some cooperation (Tikanmäki and Rathod, 2019, p. 211). Critical infrastructure, e.g. electricity and telecommunications, are an important and vulnerable part of vital societal functions. (The Security Committee, 2017; Järvenpää, 2017.).

The Finnish security cooperation model covers all levels and actors of society, as the Domestic Security Program sets intergovernmental targets for different sectors of government. The comprehensive security concept requires resource sharing, coordination, and joint planning by and between authorities (Valtonen and Branders, 2021; Tuohimaa, Tikanmäki and Rajamäki, 2011). According to the Security Committee “Preparedness measures include contingency planning, continuity management, advance preparations, training and preparedness exercises” (The Security Committee, 2017, p. 9). Global threat scenarios and disruptions that effect to the internal and external security of the society.

The description of the threat scenario in this context refers to potential disruption in the security environment that is a threat or event that jeopardizes the vital functions or strategic missions of the society. Extensive and close co-operation between authorities and other actors is needed to manage disruptions (Ministry of the Interior, 2019). National action plans for security in Finland include measures and elements that promote internal and external security. Security actors include authorities and relevant companies, legal frameworks for jurisdiction, and collaboration and information sharing for situation understanding.

On a European level, new concepts of security have shifted perceptions of internal and external threats, blurring the division between foreign and domestic policy. Internal security has an important role in operational cooperation. On an operational level, there are networks of task-based law enforcement authorities, and network management is carried out by law enforcement authorities on both macro (e.g. judges) and the micro levels (individual police and judicial authorities work together), and across borders. (Lavenex and Wichmann, 2009.)

European cooperation in space and on air, land, and maritime domains, joint capabilities, common training, and multiple collaboration projects increase safety and security. Command and Control (C2), interoperability and common strategic culture pave way to more resilient European Union (EU). The EU Global Strategy (EUGS) states: “The EU has invested significantly in the resilience of the Eastern partners, beginning with Ukraine, in areas such as rule of law, energy, critical infrastructure, cyber, strategic communications, and the reform and strengthening of the security and defence sectors”. (European Union, 2019.)

The EU has invested heavily in protecting maritime threats, such as, piracy and human trafficking. The efforts of the EU have reduced maritime accidents and helped prevent environmental disasters (European Union, 2019). The European Centre of Excellence for Countering Hybrid Threats (Hybrid CoE) improves capacity to combat and

prevent hybrid threats and enhance resilience within the EU. Hybrid CoE cooperates with NATO on hybrid and cyber issues (Hybrid CoE, 2020).

#### **4. Cooperation between authorities**

Developments in telecommunications and information systems have made our society complex and vulnerable. This development combined with threats, such as climate change, general unrest, increased violent civic activism and growing economic insecurity, as has been noted during the Covid19 epidemic. Responsibility for preparedness against crises in the Finnish society lies in the hands of e.g. the military, security authorities, the National Emergency Supply Agency, as well as security companies (Parmes, 2020.). A small country like Finland can only be effective in its crisis management when authorities and communities cooperate (Tikanmäki and Ruoslahti, 2017).

Valtonen (2010) proposes a theoretical model for cooperation between security actors, with criteria for cooperation and descriptions of cooperation processes. The author calls for effective cooperation between authorities, and this requires responsible management, confidentiality, and appropriate allocation of resources. On an individual level, a spirit and commitment to cooperate, and personal contacts become critical for successful collaborative work. Developing co-operation skills become a most important area when developing inter-authority co-operation (Valtonen, 2010.)

Krogars (2005) presents the overall process for crisis management, which lays a foundation for networking. Lanne (2007) defines a central vocabulary of security collaboration from the perspective of corporate security. The aim of her study is to develop a business security management model that could also be used in public sector security activities.

International cooperation between emergency services are hampered by differences in country-specific organizations and management systems, legislation, and security concepts. These differences make it difficult to receive international aid or to provide aid abroad. (Ministry of Interior, 2016, p. 54). Confidential relations between security authorities are essential. Extensive cooperation is needed to achieve the common goal, a safety and security community. Thus, cross-border and cross-sector cooperation is essential (McNeese et al., 2006).

The cornerstones of cooperation between public authorities and industry are situational awareness, training, and the confidential exchange of information between actors (Uusipaavalniemi and Puistola, 2016). Cyber security aims at systems and infrastructures being resilient, and situation awareness is a main prerequisite for cyber security (Pöyhönen et al., 2020).

##### **4.1 Situation picture and situation awareness**

The situation picture is a description of the common security situation and includes an analysis of the current situation and an assessment of the future. A common situation picture is an essential part of the information shared by one or more users. Common situational awareness enables collaboration task planning and assists all echelons to achieve situational awareness. (Kuusisto, 2005; Alberts et al., 2001.) The key issue in creating a situation picture is to organize the acquisition of information from different actors in society and to tailor it to the needs of each user. Creating a situation picture involves understanding the situation and assessing the evolution of the situation. "Collecting and sharing a situation picture is a prerequisite for situation management" (Tikanmäki and Ruoslahti, 2019, p. 419.)

In the User-defined Common Situation picture, the approach is network-centric and allows for multiple sources of information (Loomis et al., 2008). The Global Situation and Command and Control System (GCCS) situation picture presented by Butler et al. was intended to use existing commercial products and reduce the complexity of existing systems (Butler et al., 1996).

Smooth, seamless cross-administrative collaboration promotes situational awareness and real-time situation picture, as situational awareness (SA) is important for decision-making at different levels in various operating environments (Lehto and Limnell, 2021). The Common Operational Picture (COP) is a tool for achieving a good level of situational awareness. A good level of situational awareness requires improved decision-making abilities and precise responses as situations arise. Automation and data fusion are becoming increasingly important in

new computing platforms where humans must be able to operate. Technical systems/devices play a major role as sources of information, especially in Situation Centres' environment. (Timonen, 2018.)

In some cases, situational awareness has been considered to be a large amount of diverse information produced from multiple sources. Critical Infrastructure (CI) situational awareness has the same elements and prerequisites as traditional situational awareness, but there is a difference in the mechanisms by which the situational awareness is achieved. Command & Control (C2) systems are typically focused on geospatial thinking, while a critical infrastructure operator focuses on geographic, logical, and physical systems. (Timonen, 2018.)

Interaction and exchange of information between authorities are important for building awareness of cooperation and promoting cooperation to enhance maritime safety. Ruoslahti and Tikanmäki (2019) state that "European maritime cooperation aims at increasing situational awareness, sharing best practices, improving interoperability, removing overlapping activities, and promoting cross-border and cross-sector cooperation (p. 160).

## **4.2 Common information systems**

There are European wide and regional initiatives and developments in the field of information systems and information exchange between authorities. EU projects such as EUCISE2020, MARISA, RANGER and ANDROMEDA are producing or have made significant progress in data models and collaborative information systems (ANDROMEDA 2019; EUCISE2020 2020; MARISA 2020; RANGER 2016.). The basis for this cooperation was the European CISE Road Map in 2010, which defined the outcome of maritime information exchange and cooperation between authorities (European Commission, 2010a, b).

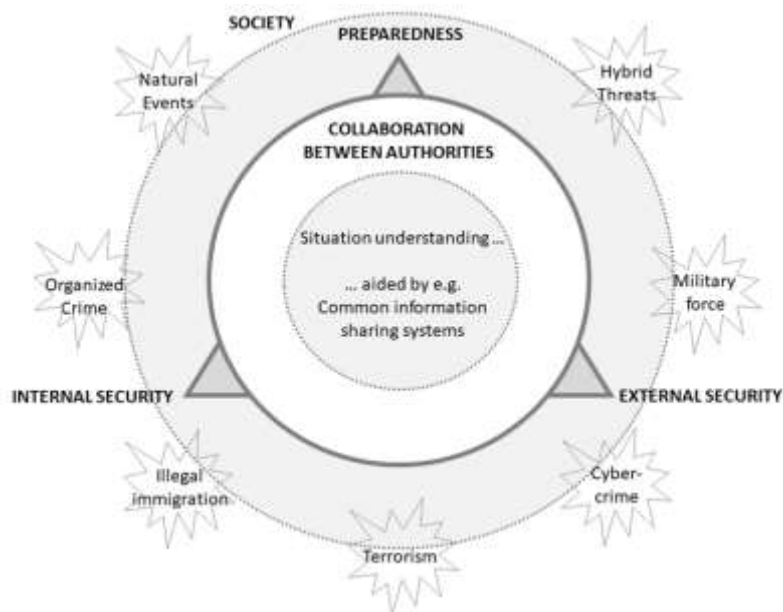
The European Commission (2013) identifies seven user communities operating on the European maritime domain: 1) Maritime Safety including Search and Rescue (SAR) and prevention of pollution caused by ships; 2) Fisheries control; 3) Marine pollution preparedness and response in Marine environment; 4) Customs; 5) Border control; 6) General law enforcement; and 7) Defence. (European Commission, 2013.) These user communities have several common IT systems in use on the land border and maritime domains.

The Finnish Police, Border Guard and Rescue Services also use the same field Command System to facilitate the coordination of accidents and other situations. A common situation picture (including available units, resources, etc.) for all authorities involved in the event promote co-operation between the rescue and police authorities, and information becomes also shared between police, rescue services, social and health services, the Border Guard, Defence Forces and Customs, and other possible authorities. (National Police Board, 2020.)

## **5. Conclusions**

In a globalized world, internal and external security cannot be separated. The distinction between internal and external security is becoming increasingly difficult as their operational environments are constantly changing and converging, as seen in Figure 1, below. In some situations, it is necessary to influence external security through internal security and vice versa. Mainly the same authorities are responsible for responding to threats, be they external or internal. The use of national, regional, and EU-wide common information sharing systems and databases to enhance cooperation between authorities are important elements in strengthening security. Rapid and up-to-date exchange of information between security authorities is needed to maintain situational awareness and understanding. Though, needed to build situation awareness, real-time information sharing may involve risks, as any collaborating entity may become subject to a cyber-breach, and the integrity of the shared data may become questionable (Pöyhönen et al., 2020).

Because cross-sectoral barriers may slow down the exchange of information between administrative sectors, staff exchanges are recommended to improve situational awareness and to develop common operating models. Increasing the knowledge levels of authorities and strengthening their exchange of information with one another across organizational boundaries, and with practitioners in need of the information promote preparedness and situational understanding (e.g. Parmes, 2020; Tikanmäki and Ruoslahti, 2017; Ministry of Interior, 2016)



**Figure 1:** Interdependence of preparedness, and internal and external security in society

State actors organized to detect and respond to hybrid threats, practice policy and operative revisions with improved cooperation between relevant actors. From the point of view of e.g. the Border Guard, hybrid threats include illegal immigration, cross-border crime, foreign fighters, and terrorism. During times of peace, cooperation with internal security actors becomes emphasized in maintaining border security (Järvenpää, 2017; The Security Committee, 2017; Ministry of Interior, 2016).

As threats change, the policies of state authorities responsible for internal and external security should be actively considered. This requires reviewing competences of these authorities and revising national legislation to enable security authorities to act when faced by threats (e.g. Tikanmäki and Rathod, 2019; Järvenpää, 2017). Since hybrid threats are both internal and external in nature, European Member States and third nations should be more willing to share information about their domestic developments. Security authorities need to 'be prepared' for anything and everything.

## References

- Alberts, D., Garstka, J., Hayes, R. and Signori, D. 2001. *Understanding Information Age Warfare*. Washington D.C.: Assistant Secretary of Defence C3I/Command Control Research Program. CCRP Publication Series.
- ANDROMEDA. 2019. "An Enhanced Common Information Sharing Environment for Border Command, Control and Coordination Systems". The European Union's H2020 research and innovation programme under grant agreement no 833881.
- Bendiek, A. 2017. A paradigm shift in the EU's Common Foreign and Security Policy: from transformation to resilience. (SWP Research Paper, 11/2017). Berlin: Stiftung Wissenschaft und Politik -SWP- Deutsches Institut für Internationale Politik und Sicherheit. Available at <<https://nbn-resolving.org/urn:nbn:de:0168-ssoar-54521-8>> [Accessed 25 September 2020]
- Butler, S., Diskin, D., Howes, N. and Jordan, K. 1996. Architectural design of a common operating environment. *IEEE Software*, vol. 13, pp. 57-65. Available at <<https://ieeexplore.ieee.org/document/542295>> [Accessed 21 March 2020]
- Cederberg, A. and Eronen, P. 2015. How can Societies be Defended against Hybrid Threats? *Strategic Security Analysis*. September 2015 No.9. Geneva Centre for Security Policy (GCSP).
- Endsley, M. 2015. Final Reflections: Situation Awareness Models and Measures. *Journal of Cognitive Engineering and Decision Making* 2015, Volume 9, Number 1, March 2015, pp. 101– 111.
- EUCISE2020. 2020. "EUropean test bed for the maritime Common Information Sharing Environment in the 2020 perspective." Available at <<http://www.eucise2020.eu/>> [Accessed 11 March 2020]
- European Commission. 2020. COM(2020) 605 final. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on the EU Security Union Strategy.
- European Commission, 2013. CISE Architecture Visions Document (Study supporting the Impact Assessment). Brussels: European Commission.

- European Commission. 2010a. COM (2010) 584 Final. Communication from the Commission to the Council and the European Parliament: on a Draft Roadmap towards establishing the Common Information Sharing Environment for the surveillance of the EU maritime domain.
- European Commission. 2010b. Integrating Maritime Surveillance. Common Information Sharing Environment (CISE). Available at <<https://op.europa.eu/en/publication-detail/-/publication/2d412889-77fd-4db5-b6fd-51b237410cf6>> [Accessed 11 March 2020]
- European Union. 2019. The European Union's global strategy. Three years on, looking forward. Available at <[https://eeas.europa.eu/topics/eu-global-strategy\\_en](https://eeas.europa.eu/topics/eu-global-strategy_en)> [Accessed 15 March 2020]
- European Union. 2018. A Europe that Protects: Countering Hybrid Threats. Available at <[https://eeas.europa.eu/topics/economic-relations-connectivity-innovation/46393/europe-protects-countering-hybrid-threats\\_en](https://eeas.europa.eu/topics/economic-relations-connectivity-innovation/46393/europe-protects-countering-hybrid-threats_en)> [Accessed 21 March 2020]
- Finlex. 2020a. Laki puolustusvoimista 11.5.2007/551. Available at <<https://www.finlex.fi/fi/laki/ajantasa/2007/20070551>> [Accessed 15 March 2020]
- Finlex. 2020b. Poliisilaki 22.7.2011/872. Available at <<https://www.finlex.fi/fi/laki/ajantasa/2011/20110872>> [Accessed 15 March 2020]
- Finlex. 2020c. Laki rikostorjunnasta Rajavartiolaitoksessa 30.1.2018/108. Available at <<https://www.finlex.fi/fi/laki/ajantasa/2018/20180108>> [Accessed 15 March 2020]
- Finnish Government. 2003. Strategy for securing the vital functions of society. In Finnish: Yhteiskunnan elintärkeiden toimintojen turvaamisen strategia. Valtioneuvoston periaatepäätös 27.11.2003.
- The Finnish Institute of International Affairs. 2018. Hybridivaikuttaminen ja demokratian resilienssi - ulkoisen häirinnän mahdollisuudet ja torjuntakyky liberaaleissa demokratioissa. FIIA Report 55/2018.
- Finnish Terminology Centre. 2017. Vocabulary of Comprehensive Security. ISBN 978-952-9794-36-2 (PDF). Available at <[http://www.tsk.fi/tiedostot/pdf/Kokonaisturvallisuuden\\_sanasto\\_2.pdf](http://www.tsk.fi/tiedostot/pdf/Kokonaisturvallisuuden_sanasto_2.pdf)> [Accessed 14 March 2020]
- Hybrid CoE. 2020. What is Hybrid CoE? Available at <<https://www.hybridcoe.fi/>> [Accessed 15 March 2020]
- Hyvönen, A. E., and Juntunen, T. 2021. From "spiritual defence" to robust resilience in the Finnish comprehensive security model. *Nordic Societal Security: Convergence and Divergence*. London: Routledge, 154-178.
- Järvenpää, M. 2017. Viranomaisten toimivaltuudet kohteiden suojaamisessa hybridiuhkia vastaan. Tiede Ja Ase, 74.
- Krogars, M. 1995. *Verkostoilla kriisinhallintaan*. Dissertation. Vaasa: Ankkurikustannus Oy.
- Kuusisto, R. 2005. From Common Operational Picture to Precision Management. Management Information Flows in Crisis Management Network. Available at <<http://julkaisut.valtioneuvosto.fi/handle/10024/78700>> [Accessed 21 March 2020]
- Lalu, P. and Puistola, J. 2015. Hybridisodankäynnin käsitteestä, Puolustusvoimien Tutkimuslaitoksen katsaus 01-2015. Helsinki: Puolustusvoimien tutkimuslaitos.
- Lanne, M. 2007. Yhteistyö yritysturvallisuuden hallinnassa. Tutkimus sisäisen yhteistyön tarpeesta ja roolista suurten organisaatioiden turvallisuustoiminnassa. Dissertation. Helsinki: Edita Prima Oy.
- Lavenex, S. and Wichmann, N. 2009. The External Governance of EU Internal Security', *Journal of European Integration*, 31:1,83 — 102.
- Lehto, M., and Limnell, J. 2021. Strategic leadership in cyber security, case Finland. *Information Security Journal: A Global Perspective*, 30(3), 139-148.
- Loomis, J., Porter, R., Hittle, A., Desai, C. and White, R. 2008. "Net-centric collaboration and situational awareness with an advanced User-Defined Operational Picture (UDOP)," in International Symposium on Collaborative Technologies and Systems (CTS), pp. 275-284.
- MARISA. 2020. "Improving maritime surveillance knowledge and capabilities through the MARISA toolkit." Available at <<https://www.marisaproject.eu/>> [Accessed 11 March 2020]
- McNeese, M.D., Pfaff, M.S., Connors, E.S., Obieta, J.F., Terrell, I.S., and Friedenber, M.A. 2006. Multiple vantage points of the common operational picture: Supporting international teamwork. In *Proceedings 50th Annual Meeting Human Factors and Ergonomics Society* (pp.467-471). Doi: 10.1177/154193120605000354
- Ministry of Interior. 2019. National risk assessment 2018. Internal Security. Publications of the Ministry of Interior 2019:9.
- Ministry of Interior. 2017. Hyvä elämä - turvallinen arki. Valtioneuvoston periaatepäätös sisäisen turvallisuuden strategiasta. Ministry of Interior publications 15/2017. Helsinki: Lönnberg Print & Promo.
- Ministry of Interior. 2016. Interdependence of Internal and External Security. Will the operational culture change with the operational environment? Available at <[http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/79230/37\\_2017\\_Interdependence%20of\\_nettiin.pdf](http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/79230/37_2017_Interdependence%20of_nettiin.pdf)> [Accessed 4 March 2020]
- Mäkelä, J. 2018. Merelliset hybridiuhat. [lecture]. Held on 23 May 2018. Finnish National Defence University.
- National Police Board. 2020. Turvallisuusviranomaisille yhteinen kenttäjärjestelmä. Available at <[https://www.poliisi.fi/poliisihallitus/tiedotteet/1/0/turvallisuusviranomaisille\\_yhteinen\\_kenttajarjestelma\\_32185](https://www.poliisi.fi/poliisihallitus/tiedotteet/1/0/turvallisuusviranomaisille_yhteinen_kenttajarjestelma_32185)> [Accessed 12 March 2020]
- Parmes R. 2020. "Varautumisen historia ja nykyhetki" in *Viestimies I/2020* pp.16-19. Newprint Oy: Raisio. ISSN 0357-2153.
- Police University College. 2020. Varautuminen eilen – varautuminen huomenna. Poliisiammattikorkeakoulun raportteja 136. Heino, O., Huotari, V. and Laitinen, K. (eds.). Tampere: PunaMusta Media Oyj, 2020.
- Prime Minister's Office. 2009. Finnish Security and Defence Policy 2009. Government Report. Prime Minister's Office Publications 13/2009. Helsinki: Helsinki University Print Bookstore.

### ***Ilkka Tikanmäki and Harri Ruoslahti***

- Puistola, J-A. 2018. Kokonaisturvallisuus ja hybridi-vaikuttaminen. [lecture]. Held on 23 May 2018. Finnish National Defence University.
- Pöyhönen, J., Rajamäki, J., Lehto, M. and Ruoslahti, H. 2020. Cyber Situational Awareness in Critical Infrastructure Protection. *Annals of Disaster Risk Sciences*, Vol 3, No 1 (2020): Special issue on cyber-security of critical infrastructure. Available at <<https://ojs.vvg.hr/index.php/adrs>> [Accessed 2 April 2021]
- RANGER. 2016. "Radars for long distance maritime surveillance and SAR operations." The European Union's H2020 research and innovation programme under grant agreement no 700478.
- Ruoslahti, H. and Tikanmäki, I. 2019. Complex Authority Network Interactions in the Common Information Sharing Environment. In *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019)*, Volume 3: KMIS, pages 159-166. September 17-19, 2019, Wien, Austria.
- The Security Committee. 2017. *The Security Strategy for Society. Yhteiskunnan turvallisuusstrategia. Valtioneuvoston periaatepäätös 2.11.2017.* ISBN: 978-951-25-2959-9.
- Tikanmäki I. and Rathod P. 2019. Enhancing the Development of Interaction between Authorities in Maritime Surveillance. In: Ntalianis K., Croitoru A. (eds) *Applied Physics, System Science and Computers II. APSAC 2017. Lecture Notes in Electrical Engineering*, vol 489. Springer, Netherlands, ISSN: 1876-1100, pp. 207-214.
- Tikanmäki, I. and Ruoslahti H. 2019. How are situation picture, situation awareness, and situation understanding discussed in recent scholarly literature? In *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019)*, Volume 3: KMIS, pages 419-426. September 17-19, 2019, Wien, Austria.
- Tikanmäki, I. and Ruoslahti H. 2017. Increasing Cooperation between the European Maritime Domain Authorities. *International Journal of Environmental Science*, Volume 2, pp. 392-399. ISSN: 2367-8941. IARAS, Nicosia, Cyprus.
- Timonen, J. 2018. *A Common Operating Picture for Dismounted Operations and Situation Room Environments.* National Defence University. Series 1: Research publications No. 19. Academic Dissertation. Tampere: Juvenes Print.
- Tuohimaa, T. and Tikanmäki, I. 2011. The Strategic Management Challenges of Developing Unmanned Aerial Vehicles in Public Safety Organizations, *10th WSEAS international conference on communications, electrical & computer engineering*, Playa Meloneras, Spain, Mar 2011, ISBN: 978-960-474-286-8, pp. 34-39.
- Tuohimaa, T., Tikanmäki, I. and Rajamäki, J. 2011. Cooperation challenges to public safety organizations on the use of unmanned aircraft systems (UAS), *International Journal of Systems Applications, Engineering and Development*, Issue 5, Volume 5, 2011 pp. 610-617.
- Valtonen, V. 2010. Turvallisuustoimijoiden yhteistyö operatiivis- taktisesta näkökulmasta. Maanpuolustuskorkeakoulu. Taktiikan laitos. Julkaisusarja 1, n:o 3. Edita Prima Oy: Helsinki.
- Valtonen, V., and Branders, M. 2021. Tracing the Finnish Comprehensive Security Model. *Nordic Societal Security: Convergence and Divergence*. London: Routledge, 91-108.



# The Host Nation Support for the International Cyber Operations

**Maija Turunen**

**Finnish National Defence University, Helsinki, Finland**

[maijaturunen@yahoo.com](mailto:maijaturunen@yahoo.com)

DOI: 10.34190/EWS.21.017

**Abstract:** International cooperation is one way to strengthen a state's cyber sovereignty, defense and create deterrence against potential adversaries. The provision or receipt of the international assistance may be considered in situations where the state is threatened or has already been the subject of an attack or military pressure contrary to the international law. A state may also allow foreign military forces to use its territory to defend a third state from an unlawful attack. The provision or receipt of the international assistance may also be considered in the situations where the humanitarian intervention is targeted at a third state. The host nation support has traditionally focused on logistical and/or material support to foreign forces on the soil of the host nation. This support may also have a cyber dimension, which makes the legal assessment of the support more complex. The nature and objectives of the operation, as well as the international law commitments of the host nation, affect the assessment of the international law and the operational challenges. In addition to legal challenges, the host nation support may also involve military, economic, political and technical challenges, such as interfaces with private actors. This paper focuses on the challenges considering the operational preparation of the environment in cyberspace. The aim is to identify, at a general level, problems that need to be resolved by the host nation before the international assistance can be provided or received. Theoretically the paper is based on the theory of the character of war, the preconditions for when the war can be waged and how the war should be waged, as well as what military actions are legitimate in war. Due to the research question, support for the cyber operations, this theory specifically applies to activities in the gray area, just before the actual escalation of the hostilities. This paper concludes that the host nation's military cyber sovereignty affects how flexibly the international assistance for cyber operations can be provided or received.

**Keywords:** host nation support, cyber operations, international assistance, international law, cyber sovereignty

---

## 1. Introduction

The provision or receipt of the international assistance may be considered in many kind of conflict situations.

The host nation may be under threat itself or it will provide support to another state. The support provided by the host nation based on either multinational or bilateral agreements. In the West, the most significant agreements are the United Nations Charter, which entitles a sovereign state to defend itself, and the partnership agreements with the North Atlantic Treaty Organization (NATO). Although the international agreements provide a framework for providing or receiving the international assistance, states decide on a case-by-case basis. The states decide to join such actions within the limits of their sovereignty. The state should be enjoys sovereignty over all cyber infrastructure within its jurisdiction. Complexity here consists of those factors (such as different dependencies) that limits state sovereignty and the difficulties of defining the cyber infrastructure under state control.

The conventional or treaty-based international law may limit a state's power to exercise its sovereign rights. From the point of view of the international law, the state sovereignty also covers a state-controlled cyber area, which may include information and communication networks and systems, equipment, digital services related to the performance of the state or its functions, and users of the above (TM2.0, 2017, 11). The cyber performance of state-owned vessels, aircrafts in the international area and satellites operating under state jurisdiction are also considered to be part of a state's cyber area (Ziolkowski, 2013, 162).

Based by United Nations Charter and other essential international agreements and judgements of the international courts, Ziolkowski (2013, 143-144) has listed the following key principles and obligations of the international law that equally belong to sovereign states: self-preservation; independence; jurisdiction over domestic matters; non-intervention in matters within the domestic jurisdiction of other States; and duty not to harm the rights of other States. Ziolkowski also argument that the principle to maintenance of international peace and security include the principles of refrain from (threat or) use of force in international relations like as duty to peaceful settlement of disputes and the duty to international cooperation in solving international problems. These rights and obligations certainly belong to all sovereign states, but their perception of the content and order of precedence of these principles varies. However, the general view is that article 51 of the United Nation Charter is a high priority: "Nothing in the present Charter shall impair the inherent right of

individual or collective self-defence if an armed attack occurs against a Member of the United Nations...” A state can retain its sovereignty and thus be able to fulfill its obligations under the international law only if it is able to organize its defense in a credible manner. This basically means the state’s own military capability and resources, but can also be based on allied relations.

The resources (technical, economical, knowledge) and political will available to the state affect how hard the state can increase its cyber sovereignty. An example of this is Russia, which has tried to strengthen its cyber sovereignty through operational, technical and legal actions, but also by building allies. Russia has given information and communication technology service providers and authorities significant responsibilities, for example to obligations to build their own sovereign information network (the RUNET) and to prepare a technical and operational readiness to isolate that from the global data network as needed at the same time (Kari, 2019, 86). The new legislation also gives more competence to the authorities, which can be seen as strengthening a military cyber sovereignty by ensuring the legislative support to use various military cyber measures also outside the country. On the other hand, Russia has also ensured significant competence for many other authorities to act in cyberspace, which could lead a resource allocation, duplication and competition between different authorities. By establishing alliances with some former Soviet republics (Collective Security Treaty Organization), Russia seeks to create a unified military information space alongside the military-geopolitical security objectives, but also exercise of the right to collective cyber defence (Kukkola, 2019, 49; Sukhankin, 2019, 322). Cooperation with the China (Shanghai Cooperation Organization) increases Russia's international influence in the cyber affairs, but can also strengthen technological development.

### **1.1 The theory of war character**

This research based on the application of the theory of war character to the cyber operations, which are mainly conducted in peacetime, *jus ad bellum*. The character of war can be defined as follows: The character of war means the common perceptions in the international system of the nature, needs and possibilities of the use of armed forces, as well as the effective principles and operating models of the armed forces. In the theory of character of war, war viewed as a pragmatic and changing phenomenon. The war character is seen in the international system and the security environment, as well as in these preconceived operational logic, strategic communication, rules, the influence of new technological advances and as a construct associated with identities of the actors (Raitasalo – Sipilä 2008, 9, Vego, 2011, 64).

In contrast to “nature of war”, which refers to those constant, universal, and inherent qualities that ultimately define war (Vego, 2011, 64), the character of war is constantly changing and tied to its creator: knowledge of what is available, the ability to interpret and utilize this knowledge. Grey (2010, pp.12-13) has stated: “War/warfare is a duel and a dynamic, unique, and unpredictable product of interaction between friendly and unfriendly forces, together with workings of friction and chance.” Vego (2011, 61) has point out that: “war is also partly in flux, constantly changing, dependent on circumstances, affected by unforeseen and incalculable events, and always requiring application through the general genius.” In this study, the formation of the character of war in the cyber environment is seen as a dynamic and fast changing system exhibits a change in response due to information, threats, will to use force and find new capabilities.

### **1.2 Methods and material**

The conceptual and technical background is a literature survey explaining the ideas and measures behind cyber operations. A context analysis of the NATO’s and United States (U.S.) guidelines and standards are used for justifying and supporting the model of war character. The primary research material consists of the official documents. Secondary, the sources include theoretical literature and academic papers on cyber warfare and cyber operations.

## **2. The host nation support for the preparation to the operational cyber environment**

The state’s decision to provide or receive international assistance is a political decision, although usually based on military considerations. The decision must also take into account the socio-economic implications and the strategic message that the decision sends to a third party, i.e. the diplomatic implications of the decision. When making a decision, the state has to assess what benefits it has from the international cooperation and what risks are involved in such activities. Through the international cooperation, a state can seek to strengthen not only its diplomatic influence in the international community, but also to improve its defense capabilities and create a

deterrent, or at least a restraint, to potential adversaries. The traditional deterrence options are deterrence by retaliation, deterrence by denial and deterrence by entanglement. Effective and active cyber defense contains elements of all of these, but in terms of collective self-defense, deterrence by entanglement in particular is close. Deterrence by entanglement is based on nation's mutual interest and political, economic, commercial, and strategic interdependence in cyberspace as well as some degree of vulnerability (Jasper, 2018, 269). The cyber deterrence can be seen to consist of the capabilities of the actor (state/adversary), the willingness to use them, and the perception that this the state manages to project to its potential adversaries. The importance of cyber deterrence will grow. As Grey (2010, 11) has stated: "The development of cyber power that is becoming ever more necessary for the creation of wealth and the functioning of armed forces already is resulting in cyber warfare. With only trivial exceptions, all future wars will harbor integral cyber warfare." However, it is difficult to see cyber deterrence as an independent entity, but rather it is part of the active and cumulative deterrence or restraint of the state.

The host nation support (HNS) is a matter of collective self-defense. Host nation support may be received from other state forces to support their own activities or to assist another state. Acting on behalf of the host nation is considered when the behavior of the operator is solely due to the host nation and when two actual conditions are met: 1) the forces are exclusively under the command and control (C2) of the host nation; and 2) the acts are committed to the purposes of the receiving state (Tallinn Manual 2.0 [TM2.0], 2017, 93). These conditions are not met, for example, in multinational NATO operations, where the command and control of the operation are usually under the NATO Operational Commander and the measures included in the operation, may also service the intentions of a non-host nation. However, these factors do not exclude the host nation's legal responsibility for its involvement in the operation.

This paper focuses on the provision of HNS to NATO operations. That based on agreements and additional protocols between NATO and the host nation. The mechanism on the ground which regulate the host nation participation in the collective cyber operations, can be described in three levels: 1) the first, general level, construct the obligations and objectives of all participating states arising from the international agreements and strategic declarations and documents of co-operation organizations; 2) the second, the state level, consists of commitments from individual host nation specific agreements, which may include, for example, memorandum of understanding on the application of general principles and practices or reservations to them. These documents may describes the concept of HNS, planning processes, policies and procedures, responsibilities of participants and practical realities of the operational implementation; and 3) the third level is the operation-specific agreements and specifications in which, the parties of the operation or other action agree with e.g. information sharing, establishing a common situational awareness, competences and management relations, technical, logistic and maintenance issues like also financial and insurance issues.

All of these things should be agreed upon and preferably exercised well in advance before the actual implementation of the operation. Cyber operations requires extensive and precise advance planning, dedication and effort (Brantly and Smeets, 2020, 5). Advance planning emphasizes the importance of sharing information. The aim is to ensure that the parties have common situational awareness, targets and understanding of the adversary's capabilities and vulnerabilities (Libicki, 2012, pp.87-91, Vázquez et al, 2012, pp.433-435).

Detailed pre-agreement and practice on the measures described above will also help to avoid legal problems during the implementation phase of the operation. Legal problems can arise when, for example, the national legislation prohibit some measures or technical equipment's or methods necessary to implement them before a state of war has been politically established. Then it is too late. Also in the cyber dimension, preparations must be made, "weapon systems" tested and troops trained before the situation escalates. Like as Libicki (2012, 37) has noted: "...norms that are inherently hard to monitor and reward cheating (e.g., against cyberweapons) or that bias cyberspace against states that believe in legislating national security behavior are far less desirable." Lewis (2015, 11) is on the same lines: "Western public opinion may demand an unrealistic level of evidence, and this could encourage opponents to attempt to evade any commitment to limit the use of cyber weapons." The Western legislation is generally built for peacetime and for civilian actors. Military needs and functions, especially in the cyber dimension and during peace, are an exception that are not often be taken into account when developing the legislation. It is therefore important for the host nation to identify at an early stage the necessary legislative changes to enable the preparation of the operational environment and the efficient and appropriate conduct of the operations also during peace, because cyber attacks against the states interests

happen all the time. Sometimes they are essential part of larger violence, the kind of hybrid warfare, as we have seen in Estonia 2007, Georgia 2008 and Ukraine 2015 (Lewis, 2015, 5).

The preparation to the operational environment, together with the host nation, aims to create and ensure a friendly cyber space for itself, as well as to prepare a cyber domain in which possible combat measures will be taken. The technical preparation of the operational environment may include, for example, changes to own or foreign information systems or the need to reserve for supported forces: networks, domains, redirectors, servers for C2, phishing and payload delivery, and so far (Huskaj –Iftimie - Wilson, 2020, 473). Problems in preparing the cyber operational environment may be related to, for example, information sharing (different states may have different regulations and standards on what kind of confidential information can be shared, to whom, under what conditions and under what technical arrangements). Cooperation with foreign forces may also require the establishment and operation of common command and communication systems or the integration of existing ones, leading to different approval procedures in different states, for example for the encryption products and software used. The third key challenge is the ownership of the networks and systems needed for the operations: in many countries, private sector actors own key information and communication networks. Therefore, the operation or deployment of the necessary information and communication networks requires cooperation with these actors. Also, the host nation's national agendas or sovereignty issues may create potential difficulties in determining cyberspace operations objectives and affect employment of cyberspace capabilities or willingness to participate in certain cyberspace operations (JP 3-16, 2019, IV-11, JP-3-12, 2013, I-3). Offensive operations in particular are often a sensitive issue for decision makers. However, the use of cyber and electromagnetic methods and technologies are essential as part of other operations (Lewis, 2015, 3, 12). Smeets (2018, 90, 92, 105) argued, that offensive cyber operations could provide significant strategic value to state actors. An offensive cyber operation can provide value in support of a national strategy they can serve as a force multiplier as well as an independent strategic asset and they can be used effectively with few casualties and achieve a form of psychological ascendancy.

The cyber operations can be roughly divided into defensive (components of defensive operations: Defensive Cyberspace Operations-Internal Defensive Measures (DCO-IDM), Defensive Cyberspace Operations-Response Actions (DCO-RA) and Defense of Non-DOD Cyberspace), offensive, and command and control system (U.S. term "DODIN") operations. Command and control system operations are constantly ongoing general process to protect own cybersecurity and operational security, while defensive and offensive operations are target-specific individual operations. (JP-3-12, 2013, pp. II-2 - II-4). In addition, cyber espionage operations are sometimes considered to be their own type of operation (Brantly and Smeets, 2020, 4). Sometimes, especially in the connection of deterrence, one can also talk about flexible defence operations (FDO) and flexible response operations (FRO). According to JP 3-0 (2017, VIII-9): "FDOs and FROs are executed on order and provide scalable options to respond to a crisis. Both provide the ability to scale up (escalate) or de-escalate based on continuous assessment of an adversary's actions and reaction. While FDOs are primarily intended to prevent the crisis from worsening and allow for de-escalation, FROs are generally punitive in nature." However, the use of cyber operations to reinforce deterrence or create restraint requires that the adversary be made aware of the target state's ability to conduct such operations. The simplest way is to communicate about this, as the U.S. and Russia did. Both admitted that they have already ability to assess to the other's power grids (Sanger and Perlroth, 2019, Nechepurenko, 2019). Both have also communicated in their strategies and doctrines the importance of developing and using cyber methods as part of their active defense.

The cyber operations are carried out in the so-called information environment, which are considered to belong cognitive dimension, informational dimension and physical dimension. These dimensions constructs a cyber space which is connected to all other domains (space, air, land and maritime). (JP 5-0, 2017, IV-11, IV-12). According to Williams (2014, 14-15) the actions to create the necessary effects in the cyber space are: cyberspace defense, cyberspace operational preparation of the environment, cyberspace intelligence, surveillance, reconnaissance and cyberspace attack. Williams (2014, 12, 15) emphasizes an ability to collect, analyze, and use intelligence information and getting an access to information and opportunities to the project power in and through cyberspace to support attaining campaign objectives and providing freedom of maneuver in cyberspace. The project power, freedom of actions/operations and defending fixed positions in cyberspace require the capability to create and protect own anti-access/area denial (A2/AD) zones, but also the ability to break adversary's A2/AD zones or create such zones in the adversary's own cyber environment. Those capabilities can prevent or inhibit an advancing force from entering an operational area (JP 3-0, 2018, I-3, I-4). JP 3-0 (2018, II-8) highlights: "Effective operational reach requires gaining and maintaining operational access in

the face of enemy A2/AD capabilities and actions. Likewise, the C2 and intelligence functions depend on operations within the EMS [electromagnetic spectrum] and cyberspace. Losing the capability to operate effectively in the EMS and cyberspace can greatly diminish the JFC's [Joint Force Commander] freedom of action."

In the cyber space, a military operation can be carried out without physically invading the adversary's soil or using the land, water or air areas of the host nation support state. Indeed, this elusiveness and secrecy are undoubtedly strengths in the use of cyber operations but at the same time create a risk of the conflict of norms in relation to the application of the international law. The exact legal nature of cyber operations on the other state's territory is somewhat unclear in the international law. According to Tallinn Manual 2.0 (2017, 20), their lawfulness can be assessed on two different bases: 1) the degree of infringement upon the target State's territorial integrity; and 2) whether there has been an interference with or usurpation of inherently governmental functions." Meeting the first criterion is challenging, as the states themselves may not have defined the cyber area under their sovereignty or the measures to protect and control it. Both criteria's are challenging in situations where the hostilities does not target the host nation itself. However, the action to prevent the spread of a conflict could be seen as part of the legitimate self-defense or collective defence.

Cyber defense needed to protect the sovereignty of the state, its core functions, its critical infrastructure and the activities and the weapon systems. Thus, the command and control system operations and defensive cyber operations are more easily reconciled within the framework of the international law. On the other hand, carrying out or participating in an offensive cyber operation poses a military and political risk in addition to legal risks, although the cyber operation characterized by deniability, impunity, and anonymity. According Tallinn Manual (2017, 104): " The wrongfulness of an act involving cyber operations is precluded in the case of: a) consent; b) self-defence; c) countermeasures; d) necessity; e) force majeure; f) distress." In assessing these criteria from the perspective of the host nation support, some observations can be made. From the host nation's point of view, the consent is in principle always fulfilled, but the scope of the consent can vary and cause problematic situations of the interpretation. The consent to purely defensive operations is easily defensible but offensive operations may also be necessary to protect the original target to be defended and for an operational security. If the support provided to prevent an attack on a third state, the preventive self-defense may become an issue if the threat is substantial, albeit not direct. However, the condition is that there is a legitimate reason for the threat or its consequences to spread to the area of the host nation.

A difficult question is that the countermeasures must be proportionate to the attack to be responded. In principle, when using a conventional armed force, the direct effects can be estimated quite accurately, but when supplying cyber-weapons, it is sometimes difficult to assess all the consequences, because the connections of the target of the attack to other objects and functions may not always be known. However, cyber operations do not directly aim at physical or human casualties but at the availability of information held by the adversary. When taking into account the principles of the international law and specially the *de minimis-rule*, it could be thought, that cyber operations are more permissible than conventional operations. It is easy to present opposing views. Geiß & Lahmann (2013, pp. 621-657), for example, raises up a number of legal issues related to cyber operations, such as the scope of the principle of self-defense, cyber-attacks as armed attacks, evidence and attribution problems, function and preconditions of countermeasures and necessity of measures.

When host nation support is support for a third country, requirements d-f are cumbersome and require a broad interpretation. It is necessary to assess, whether the operation is necessary for the target to be defended or whether the objective could be achieved by other methods. Is this a case of force majeure? Or is there an emergency situation in that third state? And what capabilities does this third state have to defend itself and conduct that cyber operation itself? However, the cyber operations take place mainly during the peacetime. Their primary purpose is to defend the interests of the state. They may also be intended to try to prevent the situation from spreading to conventional warfare or to reduce the adverse effects of the use of conventional weapons. Those problems demonstrates once again the difficult applicability of the rules of the international law adopted for the conventional warfare to cyber influence.

### **3. Conclusions**

The key conclusion of the paper is that the host nation's own military cyber sovereignty affects how flexibly it can provide or receive international assistance for cyber operations. The military sovereignty of the state linked

to the political will and legislation of the state but also to technical and economical sovereignty. The international law begins with self-defense, not preventive combating of illegal activities. The international rules of war have been created for the conventional use of force, not for cyber operations, in which case their applicability *ex analogy* depends on the applicator. The nature of the cyber operation to be supported (defensive vs. offensive vs. preparation) largely determines how the host nation support to be provided should be legally assessed. However, the host nation always takes the risk in allowing foreign state forces to use its cyber area and resources for military purposes. On the other hand, the ultimate reason for participating in collective defense and cyber operations may be the fear of being left alone in an emergency.

Offensive cyber capabilities and operations are a difficult issue for many states and especially for policy makers. However, it should not be. The fact is, that most states are developing their offensive cyber capabilities and are also willing to conduct offensive cyber operations. This is essential for a credible defense. Testing defensive capabilities requires either own or another state's offensive capabilities. States that restrict the development of their military offensive capabilities through their legislation and strategy papers are in a vulnerable position in their role as defendants in the struggle for cyberspace control.

The degree of the state sovereignty, i.e., dependencies on other states, affects how the state shapes its own character of the war. The state's military capabilities and resources as well as their distribution among the branch of defense, the state's ability to rapidly upgrade these capabilities or acquire new resources (e.g. from allies) as well as the state's willingness to use political, economic, diplomatic or ultimately military means independently affect the state's sovereignty. Thus, the states have a certain basic sovereignty but some states are able to be content with the greater enjoy a greater degree of sovereignty than others are.

Preparing a cyber operational environment takes time. Providing the host nation support for a cyber operation requires the state's preparedness, exchange of confidential information, advance planning, and exercising with potential supported partners. The good advance planning and training will also help the host nation to become aware of the legal and technical challenges of co-operation and thus develop its own legislation and the capacity and capabilities of public authorities to work together and smoothly with private sector actors. Good preparation is also a prerequisite for detecting a cyber attack quickly and effectively and countermeasure can be initiated immediately.

## References

- Brantly A., Smeets M. (2020) "Military Operations in Cyberspace." Handbook of Military Sciences. Springer, Cham.  
[https://doi.org/10.1007/978-3-030-02866-4\\_19-1](https://doi.org/10.1007/978-3-030-02866-4_19-1)
- Gray, C. S. (2010) "War—Continuity in Change, and Change in Continuity," Parameters 40, no. 2,  
<https://press.armywarcollege.edu/parameters/vol40/iss2/5>
- Geiß, R. and Lahmann, H. (2013) "Freedom and Security in Cyberspace: Shifting the Focus away from Military Responses towards Non-Forcible Countermeasures and Collective Threat-Prevention." Peacetime Regime for State Activities in Cyberspace (Tallinn, Estonia: NATO Cooperative Cyber Defence Centre of Excellence, 2013, pp. 621-657,  
<https://ccdcoe.org/uploads/2018/10/PeacetimeRegime.pdf>)
- Huskaj, G. – Iftimie, I. A. and Wilson, R. L. (2020) Designing Attack Infrastructure for Offensive Cyberspace Operations. (Proceedings of the 19th European Conference on Cyber Warfare and Security ECCWS 2020, pp. 473-482)
- Jasper, S. (2018) U.S. Strategic Cyber Deterrence Options.  
[http://centaur.reading.ac.uk/79976/1/22839264\\_Jasper\\_thesis.pdf](http://centaur.reading.ac.uk/79976/1/22839264_Jasper_thesis.pdf)
- Joint Chief of Staff (2014) JP-3-12, Cyberspace Operations. ([https://fas.org/irp/doddir/dod/jp3\\_12r.pdf](https://fas.org/irp/doddir/dod/jp3_12r.pdf))
- Joint Chief of Staff (2017) JP 5-0, Joint Operation Planning.  
[https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp5\\_0\\_20171606.pdf](https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp5_0_20171606.pdf)
- Joint Chief of Staff (2018) JP 3-0, Joint Operations.  
[https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3\\_0ch1.pdf?ver=2018-11-27-160457-910](https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_0ch1.pdf?ver=2018-11-27-160457-910)
- Joint Chief of Staff (2019) JP 3-16, Multinational Operations  
[https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3\\_16.pdf?ver=2019-11-14-170112-293](https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_16.pdf?ver=2019-11-14-170112-293)
- Kari, M. J (2019) Russian Strategic Culture in Cyberspace Theory of Strategic Culture – a tool to Explain Russia's Cyber Threat Perception and Response to Cyber Threats. JYU DISSERTATIONS 122
- Kukkola, J. (2019) "Civilian and Military Information Infrastructure and the Control of the Russian Segment of the Internet" GAME PLAYER. Facing the structural transformation of cyberspace. pp. 39-134  
<https://puolustusvoimat.fi/documents/1951253/2815786/PVTUTKL+julkaisu+11+Game+Player.pdf/a4e38a00-e30e-cc48-f3af-d590655509ba/PVTUTKL+julkaisu+11+Game+Player.pdf>
- Lewis, J. A. (2015) The Role of Offensive Cyber Operations in NATO's Collective Defence. Tallinn Paper No. 8,  
[https://www.ccdcoe.org/uploads/2018/10/TP\\_08\\_2015\\_0.pdf](https://www.ccdcoe.org/uploads/2018/10/TP_08_2015_0.pdf)

## **Maija Turunen**

- Libicki, M. C. (2012) Crisis and escalation in cyberspace. (Santa Monica: RAND Corporation)  
<https://www.rand.org/pubs/monographs/MG1215.html>
- Nechepurenko, I. (2019) Kremlin Warns of Cyberwar After Report of U.S. Hacking Into Russian Power Grid. New York Times, June 17. <https://www.nytimes.com/2019/06/17/world/europe/russia-us-cyberwar-grid.html>
- Raitasalo, J. – Sipilä, J. (2008) "Näkökulmia sotaan" Sota – teoria ja todellisuus. Näkökulmia sodan muutokseen. pp. 1-10.  
<http://urn.fi/URN:ISBN:978-951-25-1894-4>
- Sanger, D. E. and Perloth, N. (2019) U.S. Escalates Online Attacks on Russia's Power Grid. New York Times, June 15,  
<https://www.nytimes.com/2019/06/15/us/politics/trump-cyber-russia-grid.html>
- Smeets, M. (2018) The Strategic Promise of Offensive Cyber Operations. Strategic Studies Quarterly, Vol. 12, No. 3 (FALL 2018), pp. 90-113. <https://www.jstor.org/stable/10.2307/26481911>
- Sukhankin, S. (2019) "Russia's Offensive and Defensive Use of Information Security" Russia's Military Strategy and Doctrine. The Jamestown Foundation, Washington, DC February 2019, pp. 302 - 342
- Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations. Schmitt, M. N. and Vihul, L. (eds.). Cambridge 2017
- United Nations Charter (1945) <https://www.un.org/en/charter-united-nations/>
- Vázquez, D.F., Acosta, O.P., Spirito, C., Brown, S. and Reid, E. (2012) Conceptual Framework for Cyber Defense Information Sharing within Trust Relationships. pp.429-445 [https://ccdcoe.org/uploads/2012/01/6\\_5\\_VazquezEt-al\\_TrustRelationships.pdf](https://ccdcoe.org/uploads/2012/01/6_5_VazquezEt-al_TrustRelationships.pdf)
- Vego, M. (2011) On Military Theory. Issue 62, 3 d quarter 2011 / JFQ. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a546600.pdf> pp.60-67
- Williams, B. T. (2014) The Joint force commander's guide to cyberspace operations.  
([https://ndupress.ndu.edu/Portals/68/Documents/jfq/jfq-73/jfq-73\\_12-19\\_Williams.pdf](https://ndupress.ndu.edu/Portals/68/Documents/jfq/jfq-73/jfq-73_12-19_Williams.pdf))
- Ziolkowski, K. (2013) "General Principles of International Law as Applicable in Cyberspace" Peacetime Regime for State Activities in Cyberspace (Tallinn, Estonia: NATO Cooperative Cyber Defence Centre of Excellence, 2013, pp. 165-186,  
<https://ccdcoe.org/uploads/2018/10/PeacetimeRegime.pdf>)

# A GDPR Compliant SIEM Solution

Ana Vazão<sup>1</sup>, Leonel Santos<sup>1,2</sup>, Adail Oliveira<sup>1,2</sup> and Carlos Rabadão<sup>1,2</sup>

<sup>1</sup>School of Technology and Management, Polytechnic of Leiria, Portugal

<sup>2</sup>Computer Science and Communication Research Centre, Polytechnic of Leiria, Portugal

[2170101@my.ipleiria.pt](mailto:2170101@my.ipleiria.pt)

[leonel.santos@ipleiria.pt](mailto:leonel.santos@ipleiria.pt)

[adail.oliveira@ipleiria.pt](mailto:adail.oliveira@ipleiria.pt)

[carlos.rabadao@ipleiria.pt](mailto:carlos.rabadao@ipleiria.pt)

DOI: 10.34190/EWS.21.081

**Abstract:** Nowadays, cybersecurity is one of the greatest challenges that organizations are facing. One of the ways to deal with this challenge is the analysis and monitoring of computer security events to detect the numerous threats that can compromise your assets. Through the Security Information and Event Management (SIEM) systems, it is possible to carry out, in real time, the monitoring and analysis of the logs of the various systems of an IT infrastructure, and to detect and alert to possible security incidents. With the implementation of the General Data Protection Regulation (GDPR), organizations became stricter in ensuring the privacy of their employees' information, namely the data contained in the logs gathered in the various computer systems, and which contains personal data, such as IP addresses, usernames and systems accessed. Therefore, this regulation represents new challenges for the SIEM implementation. In this article, firstly the basic concepts of SIEM systems and their main functionalities were introduced. Later, the challenges posed by GDPR in the implementation of SIEM systems were also presented, namely the mandatory anonymization and pseudonymization of the sensitive data, the retention time of the logs and their encryption, and a set of technical measures that must be adopted during the implementation of a SIEM system. Afterwards, several open-source SIEM systems were compared, based on a literature review. Through this comparative study, an open-source SIEM system was elected to be used in a future implementation of a prototype, aimed to demonstrate the suitability of the technical measures previously identified as necessary for the implementation of a GDPR compliant SIEM system. In short, with this work the authors intend to identify and validate the technical measures that must be implemented in a SIEM system, in order to comply with the objectives of this type of systems and in accordance with the requirements of the GDPR.

**Keywords:** SIEM, GDPR, security incidents, log files, monitoring, legislation and regulation

---

## 1. Introduction

Currently, it is extremely challenging to map all computer security threats, since attackers have been developing increasingly sophisticated attack techniques to bypass installed security systems (ENISA, 2020). In the report of the company *Malwarebytes*, it is mentioned that in the year 2019, at the business level, there was an increase of 13% in malware compared to the year 2018 (Malwarebytes, 2020).

In this context, the centralized analysis of the security logs of applications, servers, clients and network equipment can significantly contribute to improve current security mechanisms, by identifying possible anomalies, vulnerabilities and security incidents, because it allows correlating, analysing and storing data from different sources. Usually, when we intend to centralize logs from different sources, log managers or SIEM systems are used.

The centralization and analysis of security logs can also contribute to assist organizations in ensuring compliance with GDPR, as far as logs are concerned, because it simplifies the implementation of the technical measures necessary to ensure security and privacy in the treatment of security incidents. However, when using a log manager or a SIEM system for this purpose, it is also necessary to preserve the privacy of the personal data that they store, namely using anonymization or pseudonymization techniques (Menges *et al.*, 2021). By pseudonomising the personal data contained in the security logs, it is possible for computer analysts to process this data without restrictions, since the holders of such data can no longer be directly identified. The access to the information necessary to identify the data holder is controlled by technical and organizational measures with restricted access, which ensures compliance with the GDPR. Bearing in mind what was said, the main objective of this work is to find out how or if SIEM solutions are in compliance with the requirements imposed by the GDPR. In this context, the main objectives of this work were defined: (i) the definition of the necessary requirements for the implementation of an open-source SIEM system, incorporating the necessary technical measures to guarantee the compliance with the GDPR; (ii) the selection of a SIEM solution suitable for this purpose. To achieve these objectives, it was necessary to identify the requirements that a SIEM system should



incorporate. These requirements, as well as the technical measures necessary to guarantee the compliance with the GDPR, were obtained with a literature review. The selection of the best SIEM solution was carried out through a literature review, carried out with the aim of identifying the open-source or freeware solutions on the market that are most appropriate to the intended objectives.

In summary, this article identifies the essential requirements of a SIEM system and the technical measures necessary to meet the requirements of the GDPR. A comparative study of previously selected systems is also carried out, in order to identify the solution that best suits the requirements of a SIEM system compliant with GDPR.

This paper, in addition to this section, is structured as follows. Section 2 describes the methodology adopted in this work. Section 3 summarizes SIEM and its main requirements. Section 4 introduces the main challenges posed by the GDPR and the technical measures to be implemented for a SIEM to be compliant with this regulation. Section 5 proceeds to the evaluation of the pre-selected SIEM systems and the identification of the most appropriate solution to respond to the challenges launched by the GDPR. Finally, in Section 6, the authors present a brief set of conclusions complemented with future work considerations.

## **2. Methodology**

Regarding the methodology adopted in this work, it is divided into four phases: (i) define a problem, present in the Introduction; (ii) literature review about SIEM systems and the identification of the main requirements and technical measures, necessary to meet the requirements of the GDPR, presented in section 3; (iii) define the SIEM requirements to achieve the compliance with GDPR, presented in section 4; (iv) a comparative study of pre-selected SIEM systems and the selection of the most appropriate solution to reach defined objectives, presented in section 5.

In order to achieve the objectives of this study, Google Scholar, the IEEE Xplore Digital Library and the B-on were used to analyse master's dissertations, doctoral thesis and technical reports, because this was intended to be an academic work. The Magic Quadrant (Ngo-Lam, 2020), which classifies SIEM solutions into four types (Leaders, Visionaries, Niche Players and Challengers), was also used to carry out a survey of the solutions that provide open-source or freeware versions. The results of this research enabled a more in-depth understanding of GDPR, SIEM and all the concepts related to them and to identify four SIEM systems in its open-source or freeware version: *Splunk Free* (Splunk, 2021), *Elastic Stack* (Elasticsearch, 2021), *Graylog* (Graylog, 2020), and *OSSIM* (AT&T Cybersecurity, 2021).

Later, a comparative study was carried out between four SIEM systems, in its open-source or freeware version. In this context, an individual analysis of requirements, architecture, and compliance with the GDPR was carried out for each of the systems.

After the comparative study was carried out, and based on the previously outlined requirements, the open-source SIEM system that best suits the defined requirements was selected.

## **3. Security Information and Event Management**

SIEM and centralized log management are deeply interconnected. However, they differ in their objectives, automation, and real-time analysis of security incidents. The main purpose of the centralized log management process is to collect and store data, leaving security issues in the background. On the other hand, a SIEM system, although it also ensures the centralized management of logs, has as one of its main objectives to contribute to the security of information of organizations (Catescu, 2018).

Currently, SIEM are available in various formats: software, appliances or online services (Detken, Scheuermann and Hellmann, 2015; Johnson, 2015). It should be noted that the term SIEM results from the combination of the terms Security Information Management (SIM) and Security Event Management (SEM) (Detken, Scheuermann and Hellmann, 2015).

SIEM is a tool that collects and correlates events that occurs on the network, and that allows the creation of rules and alerts that enable the detection of abnormal situations. In addition, it allows the data received (event, log or notification) from different devices (e.g., server, firewall, IDS, router) to be organized in a centralized way.

The implementation of a SIEM in an organization increases its performance in terms of security and allows the earlier identification of security incidents.

Within the scope of this work, a literature review of the essential requirements of a SIEM was carried out. For this purpose, technical documents of various solutions available on the market were consulted, as well as academic articles on this subject (Vacca, 2012; Detken, Scheuermann and Hellmann, 2015; Catescu, 2018; Graylog, 2018a; Arass and Souissi, 2019; Vazão et al., 2019; Mokalled et al., 2019; Gartner, 2020; Petters, 2020; Stefanova, 2020; Exabeam, 2021; Menges et al., 2021). Table 1 lists these requirements, accompanied by a brief description.

SIEM systems are designed to detect organization's computer security incidents. Recently, the need arose for these to be restructured, to respond to the various challenges triggered by the obligation to comply with the GDPR. In this context, it is necessary to consider that SIEM must ensure that the appropriate technical measures are applied to mitigate the risk and the severity that they can represent for the rights and freedoms of people (Vazão et al., 2019).

**Table 1:** SIEM requirements

Requirements	Description
Log management	Centralized management of logs from different components (security systems, applications, and network devices)
Log analysis	Analyses logs to identify and investigate security incidents
Log correlation	Discover and apply logical associations between events from different sources in order to identify and respond to security threats
Forensic analysis	Enables exploration of log and event data to find details of a security incident
Compliance	Ensure GDPR or Payment Card Industry Data Security Standard (PCI DSS) compliance
Application log monitoring	Every application can be logged and monitored separately as well as summarized
Real-time alerting	Alert to an anomaly or apparent security issue
User activity monitoring	Captures user actions (application usage or system commands executed)
Dashboards	Provide and create graphs where patterns and anomalies can be identified
Reporting	Provide reports on security incidents and anomalies
File integrity monitoring	Track all action types concerning files
File access auditing	Auditing accesses made to files and folders stored on computers or servers
System and device log monitoring	Track of log files and searches for known text patterns and rules that indicate anomaly or security incident
Machine learning	Uses cybersecurity rules and data to automated detection
Incident response workflows	Automate incident-response workflows
Threat intelligence feeds	Can connect to threat intelligence feeds to rapidly identify new threats

Next, the issues related to the GDPR will be introduced, as well as the challenges and technical measures that must be implemented in a GDPR compliant SIEM.

#### 4. General Data Protection Regulation

The GDPR was approved by the European Parliament on April 27 of 2016, and entered into force on May 25 of 2018, with a two-year period for its implementation in the public and private sectors (Voigt and Von dem Bussche, 2017; Saldanha, 2018; Team, 2020).

For entities that handle personal data, the GDPR differs from the previous legislation, fundamentally, in a very important point that is the value of fines for non-compliances. This amount had a significant increase, which can reach 20 million euros or 4% of the company's annual revenue (Saldanha, 2018; Team, 2020). In addition, GDPR has a much broader influence, as it includes not only European Union (EU) organizations, but all the organizations that process personal data of EU citizens or travelers in its territory (Dezeure, 2018; Team, 2020).

In this Regulation, new legal requirements are defined and it determines the way organizations should process, organize and protect personal data, with very severe financial sanctions for non-compliance, as already

mentioned (Zerlang, 2017). The purpose of this set of rules is to give citizens back control over their personal data and to regulate the business environment (Voigt and Von dem Bussche, 2017; Team, 2020). Under paragraph 1 of Article 4 of this regulation all the data that identifies its holder, or data that, while not directly identifying the person, allows easy identification of its holder (EUR-Lex, 2016; Team, 2020), is considered personal information. In other words, the concept of personal data is quite broad, as it is not limited to just the data contained in the legal identification document, covering all information in any format that allows the identification of its holder (Dezeure, 2018; Magalhães and Pereira, 2018; Team, 2020).

#### 4.1 Challenges

The GDPR changed the concept of personal information. Given that, the access, error, and security logs may contain personal information and organizations should implement technical and administrative measures to protect them (Black, 2017). It is important to keep in mind that personal data such as username, first and last name, e-mail, cookies, or IP addresses can integrate the logs that the SIEM solution stores and handles (Boucas, 2018). To minimize non-compliance with the GDPR, it is necessary to make use of techniques, such as encryption, anonymization, and pseudonymization, to reduce the risk of identifying a data holder through the personal data collected or, at least, to ensure its security when stored (Boucas, 2018; Menges et al., 2021). In addition, personal information present in the collected data should be pseudonymised, and the original information should be stored in a different system, only accessible in case of need to monitor and identify the authors of detected security incidents, namely intrusion attempts or data exfiltration attempts (Boucas, 2018; Intersoft Consulting, 2021).

For the implementation of a SIEM it is very important to consider the following articles of the GDPR: (i) article 5 - Principles relating to processing of personal data; (ii) article 15 - Right of access; (iii) article 16 - Right to rectification; (iv) article 20 - Right to data portability; (v) article 17 - Right to erasure ('right to be forgotten'); and (vi) article 32 - Security of processing. In addition, consideration should be given to the implementation of the measures necessary for the organization to follows these articles. There are other relevant articles that, in case of notification of a violation, an implementation of a SIEM system can help to address, namely: (i) article 33 - Notification of a personal data breach to the supervisory authority; (ii) article 34 - Communication of a personal data breach to the data subject; and (iii) article 58 - Powers.

#### 4.2 Technical measures for compliance

Considering the diversity of scenarios that can be found in organizations, the responsible for the processing of personal data must apply the technical and administrative measures that are appropriate to the desired level of security, considering the risk assessment (Saldanha, 2018). Cryptography and pseudonymization are two of the measures named in the following references (EUR-Lex, 2016; Magalhães and Pereira, 2018; Comissão Europeia, 2019).

To guarantee compliance with the GDPR, a list of technical measures and requirements must be ensured when implementing a SIEM. Table 2 lists the technical measures and requirements that resulted from the literature review (EUR-Lex, 2016; Black, 2017; Voigt and Von dem Bussche, 2017; Elastic, 2018; Saldanha, 2018; Splunk, 2019; Varanda, 2019; Petters, 2020).

**Table 2:** Technical measures and requirements

Technical measures and requirements	Description
Allow anonymisation of personal data	Replacement of personal data by artificial identifiers preventing the identification of the holder; the process cannot be reversed
Allow the pseudonymization of personal data	Replacement of personal data by artificial identifiers preventing the identification of the Holder; the process can be reversed
Allow retention times for personal data	Allow the definition of different retention times taking into account the categories of data
Ensure the security of personal data	Application of appropriate security measures for risk analysis
Make notifications of data breaches	Notifications using alerts for the occurrence of improper access to personal data
Restrict access to personal data	Ensure that access to personal data is strictly necessary
Audit and monitor access to personal data	Audit and monitor the operations performed by users on personal data

Technical measures and requirements	Description
Ensure resilience	Ensuring resilience or fault tolerance
Ensure disaster recovery	Ensure IT disaster recovery planning solutions
Ensure protection by design and by default	For a new service or product that processes personal data, it is necessary to design security measures in the development phase, ensuring that only strictly necessary personal data are processed
Enable the creation of Compliance Reports	Enable the creation of reports with the information required for the different compliance regulations

In short, the GDPR has added complexity to the management of information systems, as it has forced to redefine the standards and processes of an organization, in order to achieve compliance with that regulation. In the next section, based on the listed requirements, a comparative study of four selected open-source solutions will be carried out.

## 5. Evaluation and selection of SIEM GDPR compliant tools

Based on the literature analysis, it was possible to identify two open-source SIEM solutions, *Graylog* and *OSSIM*. In addition to these, it was also chosen to include in the analysis the *Elastic Stack*, because, although it is not considered a native SIEM, it allows the implementation of an open-source SIEM using other tools, as exemplified in the works of Bělousov and Marquina (Marquina, 2018; Bělousov, 2019). *Splunk Free* was also selected, since, although not open source, it offers a freeware license.

After establishing the set of the SIEM system to be studied, the researchers carried out an analysis of the requirements of the selected systems to verify, in their open-source and freeware versions, their ability to fulfill, cumulatively, requirements of GDPR and a SIEM system. This analysis will then allow us to proceed with a reasoned selection of the open-source SIEM considered most suitable for the implementation of a GDPR compliant SIEM prototype, in a production environment.

### 5.1 Analysed SIEM solutions

The requirements listed in Table 3 result from the merge of the essential requirements for the implementation of SIEM, presented in chapter 0, with the technical measures required to be compliant with the GDPR, presented in section Technical measures for compliance. Besides those, others requirements were considered relevant and were added, such as: the license, the architecture, the scalability and the authentication. It is important to note that the features indicated as unavailable in the open-source or freeware version may be available in commercial versions.

**Table 3:** Comparison between OSSIM, Elastic Stack, Splunk and Graylog

Requirements	OSSIM	Elastic Stack	Splunk Free	Graylog
License	open source	open source	freeware	open source
Architecture	appliance	on-premises/cloud	on-premises/cloud	on-premises/cloud
Scalability	No	Yes	No	Yes
Authentication	Yes	Yes	Yes	Yes
Log management	No	Yes	Yes	Yes
Log analysis	Yes	Yes	Yes	Yes
Log correlation	Yes	Yes	Yes	Yes
Forensic Analysis	Yes	Yes	Yes	Yes
Monitoring of applications/systems/devices	Yes	Yes	Yes	Yes
Alerting	Yes	No	No	Yes
User activity monitoring	No	No	No	No
Dashboards	Yes	Yes	Yes	Yes
Reporting	Yes	No	Yes	No
File integrity monitoring	Yes	Yes	Yes	No
File access auditing	No	No	No	No

Requirements	OSSIM	Elastic Stack	Splunk Free	Graylog
Machine learning	No	No	No	No
Incident response workflows	No	No	No	No
Threat intelligence feeds	Yes	No	No	No
Anonymization	No	Yes	Yes	Yes
Pseudonymization	No	Yes	Yes	Yes
Set retention time for data	No	Yes	Yes	Yes
Ensure the security of user data	Yes	Yes	Yes	Yes
Make notifications of data breaches	No	No	No	No
Restrict access to personal data	Yes	Yes	No	Yes
Audit and monitor access to personal data	No	No	No	No
Ensure resilience	No	Yes	No	Yes
Disaster recovery	No	Yes	No	Yes
Ensure protection by design and by default	No	No	No	No
Compliance reporting	Yes	No	No	No

If we choose solutions based on *cloud services*, *Splunk* and *Graylog* have limitations in terms of the volume of logs per day, the *Elastic Stack* has a limited trial period and the *OSSIM* does not offer the *cloud services* option. If we decide to implement an *on-premises* solution, the *OSSIM* solution has reduced retention limits, and *Elastic Stack* and *Graylog* have no limits.

As seen in Table 3, all the solutions allow the monitoring of applications, devices and systems, and the forensic analysis of logs. In contrast, no solution offers the functionality of *machine learning* or *incident response workflows*. According to the same table, none of the evaluated solutions provides all the features listed in the table, however the researchers highlight two features that are crucial when implementing a SIEM, scalability and log management.

Regarding GDPR, all the solutions listed in the previous table guarantee compliance with GDPR, if a paid license is obtained. However, for open-source versions, it is necessary to confirm that the features provided guarantee compliance with GDPR. Pseudonymization can be implemented by *Elastic Stack* (Wintergerst, Paquette and McDiarmid, 2018), *Graylog* (Graylog, 2018b) and *Splunk* (Varanda, 2019). For the *OSSIM* solution, considering the research carried out, no references were found that identified a possible way to implement this functionality. *Graylog*, *Elastic Stack* and *Splunk Free* solutions allow retention times to be defined. Regarding the *OSSIM* solution, due to its limit in relation to reduced retention periods, this question does not apply.

It should be noted that *Graylog* and *Elastic Stack* allow the definition of which users can access a given index. However, for the specific case of sensitive data, they do not allow an audit to be carried out on the queries made by users with permission to access the data. Due to the type of licenses made available by *OSSIM* and *Splunk Free*, only *Graylog* and *Elastic Stack* guarantee redundancy and resilience.

All four solutions guarantee a basic level of data security, ensuring the security of data in transit and restricting access to the data in the system. However, it should be noted that only commercial solutions guarantee data protection by design and by default. In addition, none of the open-source solutions offer monitoring and notification of violation of access to personal data.

## 5.2 Selection of the SIEM system

Considering the results systematized in Table 3, it is feasible to conclude that the *Elastic Stack* and *Graylog* solutions fulfil a larger number of requirements that guarantee the protection and control of personal data, in order to be in compliance with the GDPR. In addition, only *Graylog* and *Elastic Stack* guarantee scalability and resilience, which are two essential requirements in the implementation of a SIEM. Additionally, they are also the solutions that fulfil most requirements in relation to the technical measures necessary to be compliant with the GDPR.

Therefore, *Elastic Stack* was selected, due to its flexibility and because it is a solution that allows the conception of architectures that adjust more specifically to the reality of the organization, which facilitates compliance with the GDPR, as it allows the integration of other third-party open-source tools, such as *ReadonlyRest* (ReadonlyREST, 2021) or *Search Guard* (Search Guard, 2020).

Regarding GDPR, the *Elastic Stack* (from version 6.8) allows users to restrict access to indexes and features. This solution also allows the definition of various permission levels. For example, the administrators can access all data collected or produced by the SIEM in real time, but the other users only can access the information that is considered relevant to their work. On the other hand, and still in this context, the *Elastic Stack* already has a feature, named fingerprint filter, that allows the pseudonymization of data, which is one of the measures recommended by the GDPR to assure the privacy of the user's information.

Comparing the *Search Guard* and the *ReadonlyRest* plugin, the authors found that in its open-source version, a set of features is very well adjusted to GDPR compliance. In a very concise way, the authors can conclude that *ReadonlyRest* adds several levels of security to the *Elastic Stack*, because this plugin allows to perform the data access audit and to encrypt the data transported between the different components of the system: *Beats*, *Logstash*, *Elasticsearch* and *Kibana*.

## 6. Conclusions

Throughout this work, the requirements for implementing an open-source SIEM system, incorporating technical measures for the protection and control of personal data able to ensure compliance with the GDPR, were defined, and a SIEM system, to operationalize it, was selected. As a work methodology, the authors use of a literature review, and a comparative study was also carried out between four market solutions: *Splunk Free*, *Elastic Stack*, *Graylog* and *OSSIM*.

Using literature review, a literature review of the essential requirements of a SIEM and the technical measures necessary to be compliant with the GDPR was carried out. It was considered that scalability and log management are two essential characteristics for the implementation of a SIEM. Considering that the *OSSIM* and *Splunk Free* solutions are not scalable, the recommendation for the implementation of SIEM fell on *Elastic Stack* and *Graylog*. Regarding the requirements related to the GDPR, it was concluded that the pseudonymization of sensitive data and the implementation of technical measures for the protection of personal data are essential. Also, it was found that the two solutions, *Graylog* and *Elastic Stack*, provide the same requirements and allow pseudonymisation. The researchers analysed one open-source plugin, named *ReadonlyRest*, and concluded that this tool can be integrated into the *Elastic Stack* solution, adding another level of security, and allowing auditing of access to personal data. For the reasons given, the *Elastic Stack* was selected as the recommended system to implement a SIEM solution compliant with GDPR.

For future work, using the *Elastic Stack*, a SIEM prototype will be implemented in a production environment and in accordance with the GDPR, paying particular attention to the pseudo-optimization of sensitive data and the implementation of technical measures for the protection and control of personal data.

In short, due to the difficulties presented previously, the implementation of a SIEM system offers several challenges that this work aims to help overcome, especially when it is intended to implement a functional and balanced solution that is compliant with the GDPR.

## Acknowledgements

This work was supported by Portuguese national funds through the FCT - Foundation for Science and Technology, I.P., under the project UID/CEC/04524/2020.

## References

- Arass, M. and Souissi, N. (2019) 'Smart SIEM: From Big Data Logs and Events To Smart Data Alerts', *International Journal of Innovative Technology and Exploring Engineering*, Volume-8(Issue-8), p. 7.
- AT&T Cybersecurity (2021) *AlienVault OSSIM*, AT&T Cybersecurity. Available at: <https://cybersecurity.att.com/products/ossim> (Accessed: 18 January 2021).
- Bélousov, P. (2019) *Security Enhancement Deploying Siem in a Small Isp Environment*. Brno University of Technology. Available at: <http://hdl.handle.net/11012/178060>.

- Black, C. (2017) *Get One Step Closer To GDPR Compliance*. Available at: <https://www.graylog.org/resources/get-one-step-closer-to-gdpr-compliance> (Accessed: 8 March 2019).
- Boucas, E. (2018) *Importance of Using SIEM for GDPR Compliance*, *CPO Magazine*. Available at: <https://www.cpomagazine.com/cyber-security/importance-of-using-siem-for-gdpr-compliance/> (Accessed: 8 March 2019).
- Catescu, G. (2018) *Detecting insider threats using Security Information and Event Management (SIEM)*. University of Applied Sciences Technikum Wien. Available at: [shorturl.at/dtzOT](http://shorturl.at/dtzOT).
- Comissão Europeia (2019) *O que significa a proteção de dados «desde a conceção» e «por defeito»? , Comissão Europeia*. Available at: [https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-does-data-protection-design-and-default-mean\\_pt](https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-does-data-protection-design-and-default-mean_pt) (Accessed: 16 December 2019).
- Detken, K., Scheuermann, D. and Hellmann, B. (2015) 'Using extensible metadata definitions to create a vendor-independent SIEM system', in *International Conference in Swarm Intelligence*. Springer, pp. 439–453.
- Dezeure, F. (2018) 'A Layman's Guide on How to Operate Your SIEM Under the GDPR'. Splunk, p. 6. Available at: <https://bit.ly/3sa9J8s>.
- Elastic (2018) 'GDPR Compliance & The Elastic Stack'. Elastic, p. 13. Available at: <https://www.elastic.co/pdf/white-paper-elastic-gdpr-compliance-and-the-elastic-stack.pdf>.
- Elasticsearch (2021) *The Elastic Stack*, *Elasticsearch*. Available at: <https://www.elastic.co/pt/elastic-stack> (Accessed: 18 January 2021).
- ENISA (2020) *Threat Landscape 2020 - Malware*. doi: 978-92-9204-354-4.
- EUR-Lex (2016) 'Regulamento (UE) 2016/679', *Jornal Oficial da União Europeia*, (Legislação L119. 59º ano. 4 de maio). Available at: <http://eur-lex.europa.eu/legalcontent/PT/TXT/PDF/?uri=OJ:L:2016:119:FULL&from=EN>.
- Exabeam (2021) *Machine Learning for Cybersecurity: Next-Gen Protection Against Cyber Threats*, *Exabeam*. Available at: <https://www.exabeam.com/information-security/machine-learning-for-cybersecurity/>.
- Gartner (2020) *Security Information And Event Management*, *Gartner*. Available at: <https://www.gartner.com/en/information-technology/glossary/security-information-event-management> (Accessed: 30 May 2020).
- Graylog (2018a) *SIEM, Simplified | The Graylog Blog*, *Graylog*. Available at: <https://www.graylog.org/post/siem-simplified> (Accessed: 6 July 2019).
- Graylog (2018b) *What GDPR Means for Log Management | Graylog*, *Graylog*. Available at: <https://www.graylog.org/what-gdpr-means-for-log-management> (Accessed: 15 June 2019).
- Graylog (2020) *Graylog Open Source*. Available at: <https://www.graylog.org/products/open-source> (Accessed: 18 January 2020).
- Intersoft Consulting (2021) *GENERAL DATA PROTECTION REGULATION (GDPR)*, *Intersoft Consulting*. Available at: <https://gdpr-info.eu/> (Accessed: 15 January 2021).
- Johnson, L. (2015) *Security Controls Evaluation, Testing, and Assessment Handbook*. Syngress. Available at: <https://books.google.pt/books?id=X7SYBAAQBAJ>.
- Magalhães, F. and Pereira, M. (2018) *Regulamento Geral de Proteção de Dados: Manual Prático 2ª Edição Revista e Ampliada*. Edited by Vida Economica Editorial.
- Malwarebytes (2020) *2020 State of Malware Report*. Available at: [https://resources.malwarebytes.com/files/2020/02/2020\\_State-of-Malware-Report.pdf](https://resources.malwarebytes.com/files/2020/02/2020_State-of-Malware-Report.pdf).
- Marquina, L. (2018) *Ventajas e Implementación de un sistema SIEM*. Universitat Oberta de Catalunya. Available at: <http://hdl.handle.net/10609/81267>.
- Menges, F. et al. (2021) 'Towards GDPR-compliant data processing in modern SIEM systems', *Computers & Security*. Elsevier, 103, p. 102165.
- Mokalled, H. et al. (2019) 'The Applicability of a SIEM Solution: Requirements and Evaluation', in *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*. IEEE, pp. 132–137.
- Petters, J. (2020) *What is SIEM? A Beginner's Guide*, *Varonis*. Available at: <https://www.varonis.com/blog/what-is-siem/> (Accessed: 7 October 2020).
- ReadonlyREST (2021) *Documentation for ReadonlyREST plugins*, *ReadonlyREST*. Available at: <https://github.com/beshu-tech/readonlyrest-docs> (Accessed: 18 January 2021).
- Saldanha, N. (2018) *Novo Regulamento Geral de Proteção de Dados*. FCA-Edito.
- Search Guard (2020) *Licensing model*, *Search Guard*. Available at: <https://search-guard.com/licensing/> (Accessed: 20 January 2021).
- Splunk (2019) *Splunk GDPR Implementation Success*. Available at: [www.splunk.com](http://www.splunk.com) (Accessed: 11 March 2019).
- Splunk (2021) *Free vs. Enterprise*, *Splunk*. Available at: [https://www.splunk.com/pt\\_br/view/SP-CAAAE8W](https://www.splunk.com/pt_br/view/SP-CAAAE8W) (Accessed: 18 January 2021).
- Stefanova, D. (2020) *SIEM Solutions and Data Protection Compliance*, *LogSentinel*. Available at: <https://logsentinel.com/blog/siem-solutions-and-data-protection-compliance/?cookie-state-change=1610996440840> (Accessed: 18 January 2021).
- Team, I. G. P. (2020) *EU General Data Protection Regulation (GDPR)—An implementation and compliance guide*. IT Governance Ltd.
- Vacca, J. (2012) *Computer and Information Security Handbook*. 2nd edn. CRC Press.

**Ana Vazão et al.**

- Varanda, A. (2019) *O Regulamento Geral de Proteção de Dados e a Pseudonimização de Logs*. Instituto Politécnico de Leiria. Available at: <http://hdl.handle.net/10400.8/4362>.
- Vazão, A. et al. (2019) 'SIEM Open Source Solutions: A Comparative Study', in *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*. IEEE, pp. 1–5.
- Voigt, P. and Von dem Bussche, A. (2017) 'The eu general data protection regulation (gdpr)', *A Practical Guide, 1st Ed., Cham: Springer International Publishing*. Springer, 10, p. 3152676.
- Wintergerst, L., Paquette, M. and McDiarmid, D. (2018) *Protecting GDPR Personal Data with Pseudonymization | Elastic Blog, Elasticsearch*. Available at: <https://www.elastic.co/pt/blog/gdpr-personal-data-pseudonymization-part-1> (Accessed: 16 August 2019).
- Zerlang, J. (2017) 'GDPR: a milestone in convergence for cyber-security and compliance', *Network Security*. Elsevier, 2017(6), pp. 8–11.



# The Threat of Juice Jacking

Namosha Veerasamy

Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa

[nveerasamy@csir.co.za](mailto:nveerasamy@csir.co.za)

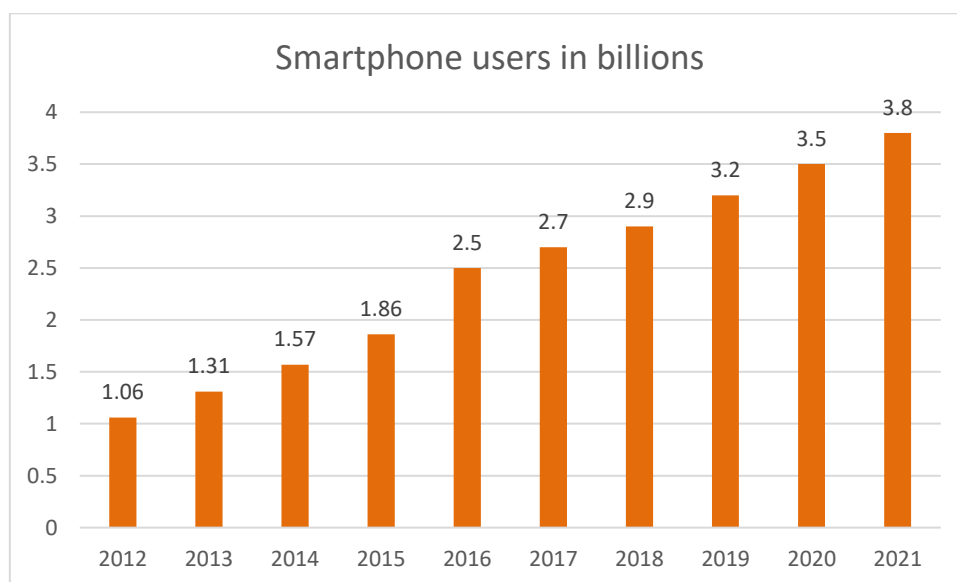
DOI: 10.34190/EWS.21.023

**Abstract:** Cyber attacks can affect the confidentiality, integrity and availability of data/ systems. Some attacks aim to steal data whereas others try cause destruction. One such vulnerability stems from the malicious use of USB chargers. When travelling and our smartphone battery level is very low, users may find a nearby charging station. However, users need to think twice before simply plugging in their device. What seems like an innocent charge could turn into a golden opportunity for attackers. Malware could actually be introduced into smartphones and other devices through the USB charger. Juice jacking is emerging as a potential risk as cyber criminals aim to infect users and potentially steal their passwords and infiltrate bank accounts. Users could even get locked out of their devices. This paper takes a closer look at this developing threat. These public charging stations are now being fraudulently used by attackers to gain access to sensitive information. Scammers are now using USB chargers as a method to steal data or install malware. However, users may be unaware of the potential risk. In this research, the malicious use of USB charging stations found in spots popular with travellers are revealed. In addition, protective measures are described in order to help users from falling victim to this latest cyber threat. Attackers try to take advantage of the situation in that most users trust their mobile devices more than their desktop devices. In addition to data theft, malicious attackers could also cause destruction of our mobile devices. When fast charging, malware could be installed onto a mobile device overwriting its firmware and arming it as a weapon. The firmware could be overwritten and the phone overloaded. The charger is thus compromised and used to overload a device. These various attack vectors are discussed in the paper to show the danger of juice jacking.

**Keywords:** juice jacking, USB, charge station

## 1. Introduction

The popularity of smartphones has grown tremendously over the past decade. The number of smartphone users worldwide today surpasses three billion and is forecast to further grow by several hundred million in the next few years- see Figure 1 (O’ Dea S, 2020) . With its increased use comes the dependency to keep the devices charged. USB charging stations offers a convenient form of keeping these devices powered on.



**Figure 1:** Smartphone users in billions (Statista .com 2020)

Many users may experience feeling of panic when their smart phone battery is about to die while on the go. The discovery of a USB charging station in public locations like airports, hotels, libraries, public transportation and malls provides some relief. However, the public needs to exercise caution before simply plugging in for a power boost. A new form of security exploitation in the form of “Juice Jacking” has emerged. Public USB charging stations can now be used to infect malware onto smartphones and other devices. Public charging stations now pose a risk and users are advised not to use them without some form of security measures. In an extreme case,

there is the potential that a free phone charge could even result in a bank account being drained due to the infection of malware that steals passwords. "A free charge could end up draining your bank account," Deputy District Attorney Luke Sisak warns, adding the malware has the ability to lock devices and share passwords with hackers (Edmond 2019). Adversaries will continue to target our smartphones with attacks like malware, malicious apps, accessibility abuse, ransomware and ad fraud. When a user's device is infected, there lies the potential for data to be read and exported, including passwords. There also lies the possibility that the device is infected with ransomware and the user is locked out of their device, making them unusable.

The concept of juice jacking was first coined by Brian Krebs in 2011 with a proof of concept at DEFCON. A free charging station was set up but when users plugged into their devices, the following warning message appeared on the kiosk (Krebs, 2011):

*"You should not trust public kiosks with your smart phone. Information can be retrieved or downloaded without your consent. Luckily for you, this station has taken the ethical route and your data is safe. Enjoy the free charge!"*

Other proof of concepts have been created over the years. Mactans was presented at the Blackhat 2013 security conference and showed a malicious USB wall charger with malware for iOS devices (Cimpanu 2019).

Then in 2016, KeySweeper a stealthy Arduino-based device was demonstrated. It was camouflaged as an operational USB wall charger but actually was able to passively sniff, decrypt, log and report back (over GSM) keystrokes of any Microsoft wireless keyboard in the vicinity (Cimpanu 2019).

Moreover in 2016, another malicious USB wall charger proof of concept was developed. This one was capable of recording and mirroring the screen of the device plugged in for the charge and lead to the concept of "video jacking" (Cimpanu 2019). Juice jacking is a type of cyberattack in which a charging port that is also used for data connection is hijacked for malicious purposes. The cyber attacker hijacks the power supply (hence juice jacking) and utilises it for offensive actions. This is done by installing malware on the victim's device and/or stealing data. The malicious programs that are installed can track the device or mirror the screen to capture passwords and PIN codes while the device is charging. This leads to juice jacking also being termed juice filming or "juice filming charging attacks" (Crane 2020). Cybercriminals wait for victims to use a USB charging connection in order to launch an attack. Hackers are aware that users may not willingly plug an unfamiliar storage device into their machine, but they think of charging cables and power banks as batteries, not IT devices (Kumar 2020). Juice jacking is comparable to card skimming scams in that an attacker sets up a malicious device over a real charging station.

USB cables can be branded to look like any other cable. This creates the impression that the cables are safe and users do not suspect that the cables are malicious. In some cases, malicious cables can even be given away as promotional gifts (Ortiz 2019). In a survey carried out by SpreadPrivacy.com in 2020 of 1029 American adults, 54.6% of respondents were not aware of the risk of public charging stations.



Figure 2: Survey of awareness of risk of public charging stations (Spreadprivacy.com 2020)

## 2. Juice jacking explained

When a smart phone is plugged into the USB port of a computer or laptop, there is an option to transfer files across the two systems. A USB port is not only a power socket but also the ability to transfer data. A standard USB connector has five pins. One pin is used to charge the receiving end. Two of the others can by default used for data transfers. Table 1 shows a summary of USB connections.

**Table 1:** USB connection table (Sunrom in Arntz 2019)

Pin	Name	Cable Colour	Description
1	V Bus	Red	+5V
2	D-	White	Data-
3	D+	Green	Data+
4	ID	N/A	Permits distinction of a host connection from device connection: Host: connected to the signal round Device: not connected
5	GND	Black	Signal ground

Unless, changes have been made in the settings, the data transfer mode is disabled by default, except on devices running older versions of Android. When a user connects to a USB port for a charge, they could be opening up a pathway through which data can be moved across.

Generally, a juice filming charging attack should have the following characteristics (Meng et al. 2019): 1) easy to implement yet efficient; 2) ease of use- i.e. user-friendly; 3) does not need the attacker/user to install additional application or component on the target device; 4) no additional permission requested from the device; 5) cannot be detected by existing anti-malware software; 6) scalable and effective on a broad range of devices ( Eg, Android, iOS, Windows); and automatic extraction of textual information from captured videos.

Juice jacking attacks can fall into two main categories:

- **Data theft:** Once the device is plugged into the compromised/fake charging station using data- transmitting USB cables, data is stolen like passwords and pins. Data theft can be automated. Malware could be planted onto the device and an additional payload dropped that steals information from connected devices. Crawlers exist that search a phone for personally identifiable information (PII), account credentials, banking related or credit card data seamlessly. Malicious apps can also clone all of the phones data onto another phone with the use of a Windows or Mac computer as an interface. An attacker can thus gain access to a wealth of information that can be used to impersonate another user. Mobile devices contain an abundance of PII that can also be sold on the dark web for profit or used as part of social engineering scams.
- **Malicious installation:** Users make use of compromised mobile device accessories like charging cables (e.g., an O.MG cable which has a hidden microchip inside the USB-C cable). Such a device appears like an ordinary lightning charging cable but it has been transformed into phone charger that can infect your device. Microcontrollers and electronic parts have become so small that attackers can hide mini-computers and malware inside the USB cable itself such as the O.MG cable (Cimpanu 2019). Attackers can make use of these exploited cables to infect the device with malicious payloads. Malware has the potential to monitor and track users' activities over a period of time. For example, malware can capture information like GPS location, purchases, social media engagements, photos, call logs and other processes. The range of malware that attackers could install includes adware, crypto-miners, ransomware, spyware and Trojans. Crypto-mining makes use of a mobile phone's CPU/GPU to mine for cryptocurrency and drain its battery. Ransomware prevents access to a phone by encrypting the device and demanding a ransom payment. Spyware results in continuous monitoring and tracking of the victim and Trojans hide in the background and can release other infections as well. Some signs that can indicate a possible infection include a slow phone, quickly drained battery, random icons on the screen, advertisement popups, notifications and a strange large phone bill. In some cases, the malware may leave no trace at all which makes it difficult to detect that the phone is infected.

An even more extreme form of juice jacking is the physical destruction of the device through digital methods. This vulnerability resides in mass-market fast chargers that are being used worldwide. When a device is connected to a fast charger with a USB cable, a negotiation occurs between the two, thereby establishing the most powerful charge that a device can handle. The management of the negotiation of this charge is handled by the firmware on the device and firmware on the charger. There is an underlying assumption that both will co-

operate with each other. However, if the charger is compromised, this negotiation can be overridden and more power can be pushed down the cable than the device is able to handle safely. This will effectively destroy the device and even potentially cause a fire. A fast charger is fundamentally a smart device and thus it can be tampered with. The attack vector consists of loading malware onto the smartphone. When the device is connected to the charger, the firmware is overwritten which makes it a weapon for whatever is plugged in next. The curveball is that the malware may be targeted at the device itself. Malware can initially be pushed onto a phone. The first time the phone is connected to the vulnerable fast charger, the phone overwrites the firmware. The next time the phone is connected to the same charger, the phone will be overloaded with power. This type of attack is termed “BadPower” and products with Badpower issues can be attacked with special hardware and target smart devices like mobile phones, tablets and laptops that support the fast charging protocol.

In research carried out by Tencet, 35 fast chargers were tested. Of those, they found “at least 18 had BadPower problems and involved eight brands.” Of those 18 charging devices, 11 were vulnerable to a simple attack through a device that also supports the fast charging protocol, such as a mobile phone (Duffman 2020). The advice offered is not to plug 5V devices with fast chargers with USB to USB-C cable. Users need to exercise care when connecting smart devices with a smart cable as it is capable of doing more than just a simple charge. These findings are indicative of the perils of the rapidly expanding IoT space. Various devices can be purchased and plugged in. Technology continues to grow with a myriad of devices and there countless little computers in the forms of phones and tablets. Data can be stolen and devices compromised. In addition, relatively innocent acts like charging a device can result in total destruction.

### **3. Protection again juice jacking**

A few steps can be taken to keep mobile phones and devices charges while on the go. The following measures can be implemented to protect against this type of threat:

- Training: Cyber awareness training to educate employees about the dangers of USB charging stations should be carried out. Users need to be educated about why they should not plug their data transmitting USB cables into public USB ports as they could potentially be exploited.
- Avoid the use of free, promotional USB charging station to prevent becoming infected
- Do not make use of plugs that are left plugged into public USB charging stations. This is comparable to the scenario whereby a lost USB is picked up from the ground. There is no way of knowing that the USB device is secure and does not contain malware and so too random technology can be tampered with and should not be implicitly be trusted.
- Only make use of USB devices from trusted reputable supplies
- If connecting, also ensure that the “Decline” option is selected when asked whether to trust the connected device
- Make use of power banks as a backup power supply. Although power banks have limited charging capabilities, they can still offer some power to hold off until a location can be found with an AC wall charger. Certain types and brands of power banks can hold enough power for several recharges. Rather invest in a high capacity power bank that can even charge multiple devices. This will help eliminate the need to look for suitable power outlets constantly.
- Make use of a USB Condom or Power-only USB cables in public. USB condoms are a devices that can serve as a buffer between the data charging cable and the public USB port. It acts as a data blocker and prevents data from being transmitted between the cable and the USB port. It limits access to the power source only and does not connect the data transfer pins. They can be attacked to the charging cable as an “always on” protection. The use of a USB data blocker or “juice- jack defender” can help prevent data exchange when the device is plugged into another device with a USB cable.
- Make use of AC adapters or power-only USB cables that can be charged through the standard AC power outlets. Carry the correct adapters for various power outlets along your route (or a universal type adapter).
- Some phones have USB preference settings. However, this is not a fully secure option. Despite setting the “no data transfer setting” data transfers have still taken place.
- Try to fully charge devices before going out.
- Non-USB options like external batteries and wireless charging can also be used.

- Switch off the device, when using a charger that is not yours. This may allow the device to be charged without any transmission of data.

#### **4. Conclusion**

Attackers are keen to find new and creative ways to infiltrate devices. Users need to remain current on the latest threats and trends. When a phone or laptop is running out of battery power, user may be keen to plug into a charging station at public locations like airports, hotels or the mall. What appears like a seemingly ordinary smartphone charge can result in a user's phone being infiltrated and infected. Through juice jacking, hackers have found an innovative way to compromise smart technology and potentially steal data or infect devices. This paper tries to create awareness on a potential exploit that can see users infecting themselves with malware or exposing their sensitive data on smart phone, tablets and other devices. The aim of this paper is to help users prevent falling victim to this type of attack and help protect themselves by describing the manner in which this attack is carried out and how to protect against this threat.

Cyber attacks are typically associated with threats like phishing, ransomware or malware. However, attackers are also keen to infiltrate devices via the USB port on smart phones. This digital form of ambushing compromises the data on a smart phone and can become a serious issue. The question arises whether this is a real threat and whether users should be concerned. From a business point of view if an attacker is able to gain backdoor into the company's data and systems, this can potentially lead to the infection of scams, malware or data theft. Ransomware or crypto miners could be planted onto a user's device. Potentially, business critical information could also be stolen. Juice jacking could potentially escalate in the future as attackers try to grow their arsenal of attacks. While it may not be as common as phishing or ransomware it is important that people are made aware of this type of threat. This paper tries show the practicality of this type of attack as attackers aim to become more ingenious. The issue could escalate in the future as hackers try to expand their attack field.

#### **5. References**

- Arntz, P. (2019) Explained: juice jacking, Malwarebytes labs. [Online]. Available at: <https://blog.malwarebytes.com/explained/2019/11/explained-juice-jacking/>, (Accessed 18 December 2020).
- Cimpanu, C. (2019). Officials warn about the dangers of using public USB charging stations. [Online], Available at: <https://www.zdnet.com/article/officials-warn-about-the-dangers-of-using-public-usb-charging-stations/>, (Accessed 18 December 2020)
- Crane, C. (2020) Juice Jacking: How Hackers steal your info when you charge devices. [Online]. Available at: <https://securityboulevard.com/2020/02/juice-jacking-how-hackers-can-steal-your-info-when-you-charge-devices/>, (Accessed 18 December 2020)
- Doffman, Z. (2020) Hackers Can Now Trick USB Chargers To Destroy Your Devices—This Is How It Works Forbes. [Online]. Available at: <https://www.forbes.com/sites/zakdoffman/2020/07/20/hackers-can-now-trick-usb-chargers-to-destroy-your-devicesthis-is-how-it-works/?sh=383c0f95bf27>, (Accessed 18 December 2020)
- Edmond, C. (2019) Hackers can use public USB chargers to steal personal data. Here's what you need to know about 'juice jacking'. [Online]. Available at: <https://www.weforum.org/agenda/2019/11/phone-cell-mobile-charging-usb-security-malware-criminal-juice-jacking-cyber-security/>, (Accessed 18 December 2020)
- Krebs, B. (2011) Beware of Juice Jacking. [Online]. Available at: <https://krebsonsecurity.com/2011/08/beware-of-juice-jacking/>, (Accessed 18 December 2020)
- Kumar, Y. (2020) Juice Jacking - The USB Charger Scam. [Online]. Available at SSRN: <https://ssrn.com/abstract=3580209> or <http://dx.doi.org/10.2139/ssrn.3580209>, (Accessed 5 January 2021).
- Meng, W. Jiang, L. Choo, K,K.R. Wang, Y. and Jiang, C. (2019) "Towards detection of juice filming charging attacks via supervised CPU usage analysis on smartphones", Computers & Electrical Engineering, Vol 78,pp 230-241.
- O'Dea, S. (2020) Number of smartphone users worldwide from 2016 to 2021 (in billions), [Online]. Available at: <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>, (Accessed 18 December 2020).
- Ortiz, A. (2019) Stop! Don't Charge Your Phone This Way. [Online] Available at : <https://www.nytimes.com/2019/11/18/technology/personaltech/usb-warning-juice-jacking.html>, (Accessed 5 January 2021)
- Spreadprivacy.com (2020) The risky business of charging your phone in public, Available at: <https://spreadprivacy.com/privacy-risks-usb-charging/>, (Accessed 6 April 2021).

# Status Detector for Fuzzing-Based Vulnerability Mining of IEC 61850 Protocol

Gábor Visky<sup>1</sup>, Arturs Lavrenovs<sup>1</sup> and Olaf Maennel<sup>2</sup>

<sup>1</sup>NATO Cooperative Cyber Defence Centre of Excellence, Tallinn, Estonia

<sup>2</sup>Department of Computer Science, Tallinn University of Technology, Tallinn, Estonia

[gabor.visky@ccdcoe.org](mailto:gabor.visky@ccdcoe.org)

[arturs.lavrenovs@ccdcoe.org](mailto:arturs.lavrenovs@ccdcoe.org)

[olaf.maennel@ttu.ee](mailto:olaf.maennel@ttu.ee)

DOI: 10.34190/EWS.21.007

**Abstract:** As smart grid technology and smart substations are becoming more common in power distribution, the use of the IEC 61850 protocol is increasing, as is the importance of the cybersecurity of the system's components. A vulnerable device can have a significant effect on the power supply in the event of a cyber-attack. The vulnerabilities of controlling devices should be identified and patched in the testing phase before deployment. Communication protocol fuzzing is a widely used, dynamic black-box testing method. It consists of sending billions of combinations of dynamically-generated incorrect input data to the device being examined and observing its behaviour. If an attempt is successful, a vulnerability can be discovered using the incorrect data as a starting point. Many IEC 61850-related vulnerability-mining research publications are available where the misbehaviour detection is based on the analysis of network traffic and the response of intelligent electronic devices (IEDs). It applies to the manufacturing message specification (MMS) protocol since it is used in the substation and determined by the client/server mode. By contrast, the generic object-oriented substation event (GOOSE) and sampled measure values (SMV) protocols are both based on a publish/subscription mechanism where no answer is expected from the clients. Therefore, other feedback solutions are needed. This paper describes a new solution. Instead of using the ping response and network traffic analysis to determine whether the device is functioning as intended, the application analyses the real-time video stream made on the human-machine interface of the tested device and the moment of the successful attempt is determined by machine learning model. Automatic video analysis can identify the input that caused an error. The paper introduces the challenges of vulnerability mining with fuzzing in GOOSE and SMV protocols focusing on the indication of the status of the tested device and characterises the developed status detector. Finally, it describes the optimal size of learning datasets and the usability and reliability of the proposed solution.

**Keywords:** fuzzing, vulnerability mining, machine learning, generic object-oriented substation event, sampled measure value

---

## 1. Introduction

The need for a decarbonised electricity supply and the increasing complexity of power distribution networks has led to radical changes. One prominent way to fulfil the requirements of energy efficiency maximisation beside rapid reaction capability is to replace ageing assets with modern information and communication technologies (ICTs). Smart Grid uses advanced technologies to control the power distribution reliably and efficiently (Yokoyama et al., 2013). The IEC 61850-based smart substations have played a significant role in power system operation, becoming increasingly complex and interconnected as state-of-the-art ICTs are adopted. The increased complexity and interconnection of supervisory control and data acquisition (SCADA) systems have exposed them to a wide range of cybersecurity threats. The reliability and stability that is based on the errorless implementation of the controlling software of the used intelligent devices are crucial regarding the security of the power supply. The impact of the exploitation of a vulnerability is unpredictable, not to mention the cost of the patching. The vendors should test their devices before putting them on the market to find possible errors in implementation.

Fuzz testing is one of the most useful techniques in finding vulnerabilities of software and protocol implementations (Sutton et al., 2007). In protocol fuzzing, a fuzzer sends virtually unlimited test cases using invalid or manipulated data to protocol implementation, within the framework defined by the protocol specification. Using effective test cases as input information enables security vulnerabilities to be found which were not anticipated by the protocol designers or software developers and is an effective approach to improving security and reliability (Ai-Fen Sui et al., 2011).

In conventional software fuzzing, there are several methods used to monitor the behaviour and health status of the application, like process monitoring or network traffic analysis. But in the case of industrial control systems,

we gain little data on the internal status of the controller therefore network traffic analysis can be used. This feedback method is applicable in the case of modern substations if the manufacturing message specification (MMS) protocol is applied. MMS is a client-server protocol, so the fuzzer should expect a response from the controller. The generic object-oriented substation event (GOOSE) and sampled measure value (SMV) protocols are both based on a publish/subscription mechanism where no answer is expected from the clients and therefore other feedback acquisition solutions are needed.

Yang et al. (2015) stated that one of the common outputs of successful fuzzing is that the human-machine interface (HMI) of the tested IED has no response, or it shows an error state. Assuming that the exact moment of the successful fuzzing is logged, it can be correlated to the log of the packets sent to the device. This method could make vulnerability mining more effective.

This paper's main contribution is pinpointing the exact moment at which the HMI of the inspected device starts displaying an error state, indicating the abnormal condition of the controller. To accomplish this, we created a deep neural network based on TensorFlow framework that can classify device state based on display image input. The proposed novel application of ML in a legacy industrial sector might help to accelerate innovation and overcome resistance to new solutions, at the very least in the security context.

In Section 2 we review related research addressing fuzzing industrial protocols and the applicability of ML for fuzzing. In Section 3 we provide background on IEC 61850 protocol. In Section 4 we describe the testing setup and discuss limitations. Section 5 provides the results, and Section 6 the conclusions.

## **2. Related work**

Fuzzing is a highly researched topic in software testing, primarily in the context of vulnerability identification. A systematic review of this research field is provided by Chen et al. (2018) and Liang et al. (2018). While fuzzing can be applied in a wide range of environments, we are focusing only on the network protocol fuzzing which is the only way to interface with the ICS devices.

There are several studies published regarding the creation of fuzzing data (Ai-Fen Sui et al., 2011; Blumbergs and Vaarandi, 2017) and testbeds (Yang et al., 2015; Mathur and Tippenhauer, 2016) where the ICS devices can be tested, but they rarely explore the indication of the corruption of the inspected device. Tu et al. (2017) introduced the IECFuzzer containing four modules: the mutator module, the mutator selection module, the data reorganisation module and the survival verification module. Here the survival verifier module communicates with the target PLC in real-time and obtains the status of the PLC. This solution supposes the existence of a communications channel that works even if the inspected device is in a broken state. Our solution focuses on the HMI that indicates error status visually.

The application of machine learning (ML) and AI for network fuzzing has emerged as a way of simplifying the testing of complex protocols and decreasing the time needed to set up testing (Saavedra et al., 2019). The most common application of ML in fuzzing is generating inputs based on samples instead of defining grammar manually. These inputs can be either injected directly into the software such as in a PDF file (Godefroid et al., 2017) or into the network as packets. A generative adversarial network has been proposed to learn and generate industrial network protocol message (Hu et al., 2018). A deep neural network predicting the likely vulnerability path to prioritise fuzzing mutation energy is explored by Wang et al. (2019). These studies differ from our proposal as we are applying ML to determine the fuzzed device state instead of inputs.

Machine Vision (MV) is an independent research field capable of image classification. It can recognise people, objects, places, actions and writing in images. The combination of AI software and MV technologies can achieve an outstanding result in image classification. Recently, image classification has been growing and becoming a trend among technology developers especially in e-commerce and automotive, healthcare and gaming industries (Rastegari et al., 2016). The task of image classification is to make sure all the images are categorised according to their specific sectors or groups. Image classification is easy for people but has proven to be a major problem for machines. The problem comes when unidentified patterns are compared with known ones (Xie et al., 2015). For the investigated research problem, MV alone is insufficient for the classification as completely valid display values might still indicate an error state. But the MV could provide optimisation by producing processed output for the feeding into ML classifier.

In recent years, a variety of software libraries have been released which accelerate the research and application of neural networks. TensorFlow, created by Google, is the most popular (Pang et al., 2020). It is a flexible and scalable software library for numerical computations using dataflow graphs. This library and related tools enable users to efficiently program and train neural networks and other machine learning models and deploy them to production (ibid.)

### 3. Background

#### 3.1 Introduction of IEC 61850

The international communication standard for devices within a substation environment is International Electrotechnical Commission (IEC) 61850 and it has contributed immensely to the way communication and information exchange are implemented within an electrical substation (Kriger et al., 2013). The main goal of the IEC 61850 standard is to improve the automation of electrical substations. The main objectives are to avoid proprietary protocols and provide the ability to integrate equipment from different manufacturers using technologies that can reduce the cost in wiring and engineering time, and seeking improvements in commissioning and maintenance.

IEC 61850 defines an abstract data model that can be mapped to multiple protocols. Common protocols include the MMS, GOOSE and SMV protocols (Adamiak, 2004). Figure 1. is a comparison between the IEC61850 protocol stack and the five-layer protocol stack of the common computer network (Falvo et al., 2013).

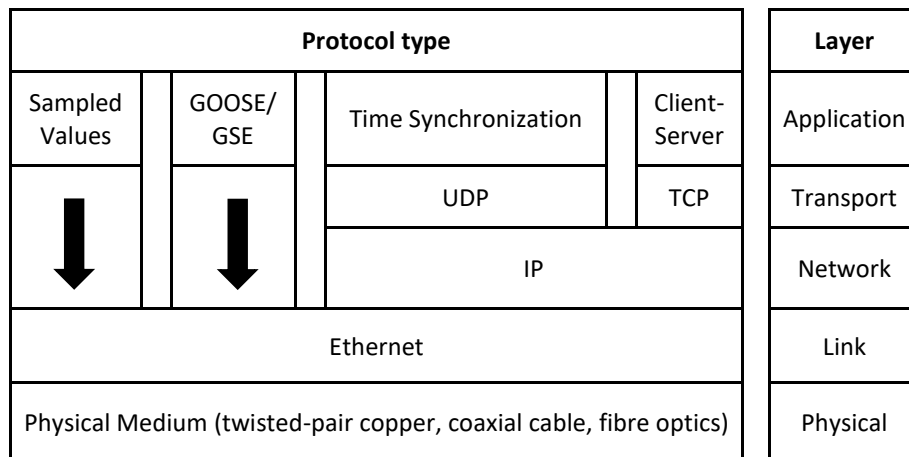


Figure 1: Comparison between IEC61850 and 5-layer protocol stack

The communication protocols of IEC61850 can be divided into two main groups: the client-server communication that takes place between the servers (the protection and control devices) and the clients. Here, SCADA and devices can act as clients. This kind of communication is implemented with an MMS protocol defined in ISO 9506. These messages are routable and sent over layer 3 of the Open System Interconnect (OSI) model and used for control and automation functions (El Hariri et al., 2016). A TCP channel is created between each client and each server. Over this channel the client can read data, force settings, request commands or receive spontaneous reporting.

IEC 61850 defines fast and reliable point to multi-point message exchange procedures that can be used for data exchange between the sensors and controllers in a substation.

These are the GOOSE and SMV protocols. GOOSE messaging is a fast, non-routable, reliable data exchange protocol between IEDs defined in IEC 61850-8-1, which is the basis for critical power system functions such as status changes, blockings, releases or power line protection, while SMV is used to quickly transmit a synchronised stream of current and voltage data samples. These are Ethernet messages sent over layer 2 of the OSI model (IEEE 802.3) following a publish/subscribe model, meaning, the publisher of the data has no feedback from the subscribers, so this should be considered unidirectional point-to-point or bidirectional multi-point communication.



### **3.2 Overview of fuzzing**

We expect high reliability from basic operating systems, at the same time the formal verification of a complete set of system utilities is too onerous a task (Miller et al., 1990). There is still a need for some form of more complete testing since the addition of randomness to the input could trigger bugs that were not found during the software's testing processes. These bugs are usually based on the lack of input value verification and lead to buffer-, stack- or arithmetical overflows, race conditions and return value failures. These bugs are often the sources of security vulnerabilities (Saito et al., 2016). Modern ICTs, especially industrial devices, have highly sophisticated input validation and can handle random input data, but there is still some chance of mistakes in implementation that can be located and exploited.

Fuzzing is a black-box testing method in which random input data is sent to the tested device and its behaviour monitored. Historically, success is declared when a fuzzer reveals a bug that is harbouring a vulnerability. However, for critical infrastructure, we are considering a broader definition of success: discovering a software bug that creates any sort of disruption. We care about all disruptions as any disruption, whether a security vulnerability or not, may affect the stability of the system.

When an application is being fuzzed to detect a vulnerability, the most important part of this process is to monitor it. Debuggers are typically used to monitor the target application during the test. Besides debugging the application, other parameters should be monitored during the fuzzing process including memory use, file system activities and registry file access. These pieces of information are rarely available in ICS devices since there is usually no dedicated data source implemented for monitoring purposes.

Monitoring network activity is a possible solution that could be applied in the industrial environment. Sending input data and analysing the replies could lead to error detection. This could be used with protocols where a response is expected from the client, like the MMS. With multicast protocols like GOOSE or SMV, this solution would not work as there is no answer from the client. This can be solved by sending a command or data to the inspected device while continuously checking its status. For example, if a device does not answer a ping request, we might suppose that it has been corrupted. Supposing that the HMI of the controller reflects a broken state on its display, this information could be used to identify the data that caused the malfunction. This approach needs a detector that can point out the moment of failure and that can be correlated with the fuzzing logs.

## **4. Setup for corruption detection**

To pinpoint that moment, a detector system is including a camera which captures the video stream of the HMI display and a computer that continuously analyses the video stream and writes the results into the status log.

There are three main parts of the procedure: preparation, model creation and training, and fuzzing. As in any deep-learning framework, one of the most important steps is the preparation of a dataset for the training and testing phase, the dataset plays a vital role since the deep-learning convolutional neural network needs a lot of training, testing and validation samples (Janahiraman and Subuhan, 2019). During the preparation phase, two sets of footage are recorded: first, where the controller functions properly and the display shows the values that could be present in the case of normal operation; second, where the controller is in error mode and the HMI displays error code or abnormal values. The footage is recorded on the HMI, so it contains displays but not static components such as buttons or labels. To create the training data set, the footage is processed. In the preparation phase, the picture frames with different content are extracted from both sets. Depending on the speed of the camera, every 10<sup>th</sup> to 25<sup>th</sup> frame is stored as a picture. The images extracted from the two sets are used to train the data model after the image resolution has been changed to 224x224 pixels. OpenCV library is used to transform the images.

After the preparation of the training data the model training comes. In our solution, we used the MobileNetV2 TensorFlow image classification model, which is a general architecture of broad application with varying input layer sizes. This allows different models to reduce computation and thereby reduce inference on mobile devices.

The third part of the procedure is the near real-time classification of the pictures. During this phase, the video stream is processed and converted into pictures, frame-by-frame, just like during the preparation, including the resolution modification. The pictures extracted from the video stream are classified by the ML-based classifier and the labels of the pictures logged. To keep the amount of data to a minimum, the speed of the recording can

be reduced to two images from each displayed value on the HMI. This can detect the moment when the status of the tested device changes from normal to error. Source code is available in author's git repository (<https://gitlab.com/viskyg/video-classifier>).

#### 4.1 Research environment

We used a self developed computer software that mimics a part of the HMI of the tested controller with four of seven-segment 5-digit LED displays. In normal operation mode, as it can be seen in Figure 2, the values of the four different displays legitimate values, that continuously fluctuate with a 1 second refresh time in the simulator application.



Figure 2: HMI in normal mode

If the controller is broken because of a successful attempt, the display of the simulated controller shows the broken state as in Figure 3. One display shows 'Err' representing the error status, one display shows a significantly different value (0). The other two displays function just like in the normal condition, so the digits are continuously changing.



Figure 3: HMI in error mode

This data source fulfils the requirements of our research since the screenshots can be used to train the detector during the preparation then the screenshots of the different statuses can be labelled. In a real environment, we would operate on the same data but from a different source, so this simulated HMI provides circumstances close enough to the future use environment.

The detector application can label the extracted pictures and point out the actual status of the controller in real-time. The log of the status combined with the log of the fuzzer helps to determine the moment of the successful attempt for further investigation. Source code is available in author's git repository (<https://gitlab.com/viskyg/hmi-display>).

#### 4.2 Challenges for physical displays

The algorithm performed well with the artificially generated pictures, but in the real environment, there were some additional challenges. With a physical display, the type of display can influence the results for several reasons. In some cases, refreshing displays built from industry-standard seven-segment elements can interfere with the camera, so individual pictures might contain only a couple of active segments. The real displayed values can be determined by averaging the pictures. We faced difficulties with the backlights of LCDs, since when inactive they might switch off, so before fuzzing the 'switch off time' must be set to infinite or the backlight

switched off in favour of artificial lighting. The light source and camera must be at a proper angle to get the best picture. There might be further challenges with more sophisticated displays when the refresh rate of the display interferes with the speed of the camera. All these issues might require precise camera and lighting tuning combined with image post-processing.

The speed of classification and refresh period of the display can also limit the accuracy of detection. Our solution on average classified 21 picture frames per second, therefore the refresh rate of the display can have a significant influence on accuracy.

### 5. Results

We evaluated the influence of the training data on picture classification. As can be seen in Figure 4, the number of training pictures did not significantly influence the results of the classification. The classification error came from the model training and happened regardless of the training.

Training sample size		Correct result		Wrong Result	
Error	Ok	True positive	True negative	False-positive	False-negative
1800.00	1800.00	0.50	0.50	0.00	0.00
900.00	900.00	0.50	0.50	0.00	0.00
450.00	450.00	0.50	0.50	0.00	0.00
350.00	350.00	0.46	0.50	0.00	0.04
300.00	300.00	0.50	0.50	0.00	0.00
250.00	250.00	0.50	0.50	0.00	0.00
200.00	200.00	0.50	0.50	0.00	0.00
180.00	180.00	0.50	0.42	0.08	0.00
150.00	150.00	0.50	0.50	0.00	0.00
120.00	120.00	0.50	0.50	0.00	0.00
80.00	80.00	0.47	0.50	0.00	0.03

Figure 4: Results with different training sample sizes

To define the optimal number of training frames we use confidence intervals. Confidence intervals are a way of quantifying the uncertainty of an estimate. We suppose that the labelling algorithm labels the good cases as good with a probability of 1 and as bad (1-1) with a probability of 0. In real-life, these probabilities will be closer together, so we propose to use the smallest training dataset when the difference is greatest and thus reliability of labelling is highest. As can be seen in Figure 5, the best performance can be expected if 120 bad and 120 good pictures are used for training.

Sample Number	1800	900	450	350	300	250	200	180	150	120	80
Probability Difference	0.49	0.40	0.44	0.36	0.35	0.39	0.71	0.65	0.67	0.74	0.69

Figure 5: Probability differences

In Figure 6 the graphical representation of the probability difference can be seen.

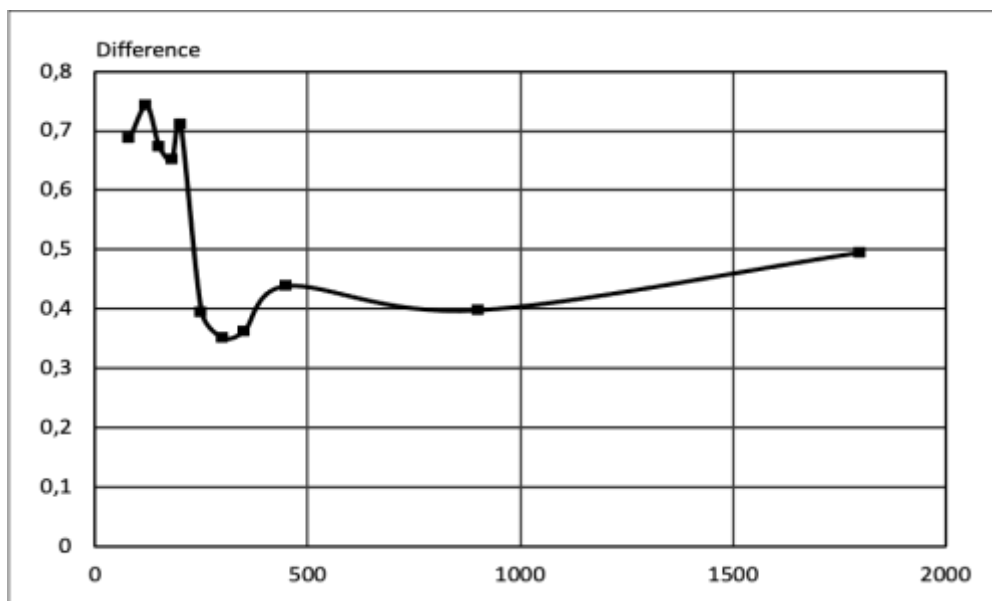


Figure 6: Probability differences

## 6. Conclusions

Fuzzing complex industrial network protocols is an efficient way of locating vulnerabilities. The application of AI for input generation has significantly increased coverage and decreased the time required for testing. However, because of the nature of ICS devices and industrial network protocols, there are circumstances in which it is not possible to directly establish a debugging connection with the tested device to determine its operational state and the success of the fuzzing. In this paper, we have proposed a status detection solution based on ML that can be used during IEC61850 SMV and GOOSE protocol fuzzing. The developed status detector is based on image classification using deep-learning via a TensorFlow framework. The moment when the HMI starts displaying the error status is detected, and therefore the human workload required for testing can be reduced significantly.

While the proposed solution was tested on generated input images, the approach was the same for physical displays but all the physical aspects of picture capture must be considered including refresh rate of both camera and display, lighting, angles, backlight and light spots. If higher performance is needed, then MV can be used for more efficient recognition of well-defined displays and model trained on the processed numerical and character inputs.

## References

- Adamiak, M. (2004) IEC61850 Communication Networks and Systems in Substations: An Overview, p. 16.
- Ai-Fen Sui et al. (2011) An effective fuzz input generation method for protocol testing, in *2011 IEEE 13th International Conference on Communication Technology. 2011 IEEE 13th International Conference on Communication Technology (ICCT)*, Jinan, China: IEEE, pp. 728–731.
- Blumbers, B. and Vaarandi, R. (2017) Bbuzz: A bit-aware fuzzing framework for network protocol systematic reverse engineering and analysis, in *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM). 2017 IEEE Military Communications Conference (MILCOM)*, Baltimore, MD: IEEE, pp. 707–712.
- Chen, C. et al. (2018) A systematic review of fuzzing techniques, *Computers & Security*, 75, pp. 118–137.
- El Hariri, M., Youssef, T. and Mohammed, O. (2016) On the Implementation of the IEC 61850 Standard: Will Different Manufacturer Devices Behave Similarly under Identical Conditions?, *Electronics*, 5(4), p. 85.
- Falvo, M. C. et al. (2013) Technologies for smart grids: A brief review, in *2013 12th International Conference on Environment and Electrical Engineering. 2013 12th International Conference on Environment and Electrical Engineering (EEEIC)*, Wroclaw: IEEE, pp. 369–375.
- Godefroid, P., Peleg, H. and Singh, R. (2017) Learn&Fuzz: Machine learning for input fuzzing, in *2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE). 2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE)*, Urbana, IL: IEEE, pp. 50–59.
- Hu, Z. et al. (2018) GANFuzz: a GAN-based industrial network protocol fuzzing framework, in *Proceedings of the 15th ACM International Conference on Computing Frontiers. CF '18: Computing Frontiers Conference*, Ischia Italy: ACM, pp. 138–145.
- International Electrotechnical Commission (1995) IEC 61850 Communication Networks and Systems In Substations.

- Janahiraman, T. V. and Subuhan, M. S. M. (2019) Traffic Light Detection Using Tensorflow Object Detection Framework, in *2019 IEEE 9th International Conference on System Engineering and Technology (ICSET)*. *2019 IEEE 9th International Conference on System Engineering and Technology (ICSET)*, Shah Alam, Malaysia: IEEE, pp. 108–113.
- Kruger, C., Behardien, S. and Retonda-Modiya, J.-C. (2013) A Detailed Analysis of the Generic Object-Oriented Substation Event Message Structure in an IEC 61850 Standard-Based Substation Automation System, *International Journal of Computers Communications & Control*, 8(5), p. 708.
- Liang, H. et al. (2018) Fuzzing: State of the Art, *IEEE Transactions on Reliability*, 67(3), pp. 1199–1218.
- Mathur, A. P. and Tippenhauer, N. O. (2016) SWaT: a water treatment testbed for research and training on ICS security, in *2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)*. *2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)*, Vienna, Austria: IEEE, pp. 31–36.
- Miller, B. P., Fredriksen, L. and So, B. (1990) An empirical study of the reliability of UNIX utilities, *Communications of the ACM*, 33(12), pp. 32–44.
- Pang, B., Nijkamp, E. and Wu, Y. N. (2020) Deep Learning With TensorFlow: A Review, *Journal of Educational and Behavioral Statistics*, 45(2), pp. 227–248.
- Rastegari, M. et al. (2016) XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks, in Leibe, B. et al. (eds) *Computer Vision – ECCV 2016*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 525–542.
- Saavedra, G. J. et al. (2019) A review of machine learning applications in fuzzing, *arXiv preprint arXiv:1906.11133*.
- Saito, T. et al. (2016) A Survey of Prevention/Mitigation against Memory Corruption Attacks, in *2016 19th International Conference on Network-Based Information Systems (NBIS)*. *2016 19th International Conference on Network-Based Information Systems (NBIS)*, Ostrava, Czech Republic: IEEE, pp. 500–505.
- Sutton, M., Greene, A. and Amini, P. (2007) *Fuzzing: brute force vulnerability discovery*. Upper Saddle River, NJ: Addison-Wesley.
- Tu, T. et al. (2017) A Vulnerability Mining System Based on Fuzzing for IEC 61850 Protocol, in *Proceedings of the 2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017)*. *2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017)*, Taiyuan, China: Atlantis Press.
- Wang, Y. et al. (2019) NeuFuzz: Efficient Fuzzing With Deep Neural Network, *IEEE Access*, 7, pp. 36340–36352.
- Xie, L. et al. (2015) Image Classification and Retrieval are ONE, in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. ICMR '15: International Conference on Multimedia Retrieval*, Shanghai China: ACM, pp. 3–10.
- Yang, Y. et al. (2015) Cybersecurity testbed for IEC 61850 based smart substations, in *2015 IEEE Power & Energy Society General Meeting. 2015 IEEE Power & Energy Society General Meeting*, Denver, CO, USA: IEEE, pp. 1–5.
- Yokoyama, A. et al. (2013) *Smart grid: technology and applications*. Hoboken, N.J.: Wiley.

# Mobile Phone Surveillance: An Overview of Privacy and Security Legal Risks

Murdoch Watney

University of Johannesburg, South Africa

[mwatney@uj.ac.za](mailto:mwatney@uj.ac.za)

DOI: 10.34190/EWS.21.021

**Abstract:** The discussion focuses on the collection, use and disclosure of personal information pertaining to a mobile phone and the circumstances in which state and corporate (non-state) surveillance may not be lawful. It highlights the tension between government, law enforcement agencies, companies, businesses and the users of mobile phones relating to mobile phone surveillance. It revolves around data control. It also touches on the ownership of information on a mobile phone and the apps downloaded on a smart phone. It emphasises that personal information has substantial value. There are many stakeholders who want access to this information. Law enforcement may wish access to it for investigating a crime whereas companies may want access to it to profit from it by means of advertisement revenue, for example. The issue of phone surveillance came under serious global scrutiny in 2013 when a United States (U.S.) National Security Agency (NSA) contractor, Snowden, disclosed that the NSA had secretly been building a vast database of US telephone records. The disclosure of the government's violation of privacy impacted negatively on government accountability and public trust. Seven years later in 2020, the U.S. Supreme Court of Appeals for the Ninth Circuit found that such warrantless bulk surveillance had been unconstitutional. Mobile phones were unfortunately not designed with the emphasis on privacy and security. Although state and non-state surveillance must take place within legislative parameters and should be subjected to checks and balances, the circumstances in which access to mobile phone information for various purposes may be gained and how it may be obtained, should be scrutinised regularly. Post Snowden the focus was mainly on state surveillance, but currently the enormity of the threat of surveillance capitalism is being appreciated. Corporate-enhanced abilities to acquire, manipulate and sell personal information may seriously undermine privacy protection. This discussion provides an overview from a legal perspective of the various aspects pertaining to state and capitalistic (non-state/corporate) surveillance which may pose security and privacy risks to mobile phones users.

**Keywords:** mobile phone surveillance, collection, use and disclosure of mobile phone data, privacy and security risks of mobile phone usage, state surveillance, surveillance capitalism

---

## 1. Introduction

It is estimated that the current number of users of mobile phones (including both smart and feature phones) approaches 4.88 billion users, which makes 63.60% of people in the world a mobile phone owner (Bankmycell, 2021). Feature phones are cell phones that have basic capabilities such as the ability to make phone calls, send SMS text messages and provide access to the Internet. The current number of smartphone users in the world is 3.8 billion which means 48.53% of the world's population owns a smartphone (Bankmycell, 2021). South Africa's smartphone penetration reached 91.2% in 2019, up from 81.7% in 2018 (Mzemkandaba, 2020). Africa is known to be a mobile-first continent where the majority of Internet access is achieved through mobile devices.

Mobile phones have become an essential part of human life. This is reflected in the number of people that own a mobile phone. It is undoubtedly one of the most important electronic devices today. It has drastically changed the manner in which people live and work. The 2020 COVID19 pandemic contributed to the growth of the smart phone market as it compelled people to go online for working and/or studying purposes.

Having a mobile phone in one's pocket simplifies the lives of users. Smartphones provide for emailing, downloading music and various apps, playing games, accessing messaging apps such as WhatsApp and social media services, such as Facebook and Instagram anywhere and at any time. For example accessing the Internet of Things (IoT) by means of a smart phone, makes it possible to interlink a myriad of gadgets and exchange information quickly and easily.

The advantages of the use of mobile phones, specifically smart phones, are therefore numerous. The downside is that mobile phones were not designed with the emphasis on security and privacy.

The cell phone was designed to be a surveillance technology – the network must know a phone's location in order to route calls to the phone (Wicker, 2013). Smartphones have therefore become a human "electronic tag" that users constantly carry with them. Precise locations, dates, times, durations and what users did before and

after opening an app or website are all tracked by the phone carrier and the companies that provide their favourite services.

Connectedness has opened the door to surveillance. Cyber criminals already access data like bank logins, credit card numbers and more. Strong security is essential. Many of the IoT devices are, and will continue to be, accessed via smartphones. While this is very convenient for users, there are weaknesses in smartphone security that may be exploited to turn smart objects against users.

Various countries have legislation that provides for the protection of personal information, thus ensuring consumer privacy protection. Countries also have legislation which provides for circumstances in which surveillance for law enforcement purposes may lawfully take place. Despite legislative provisions regulating access to mobile phone information, unchecked surveillance still presents security and privacy risks.

This discussion focuses on the surveillance of mobile phone data for various purposes and it is discussed from a legal perspective. The following aspects will be considered:

- Part 1: Law enforcement access to mobile phone information;
- Part 2: Access to consumer' information by companies, businesses and social media messaging apps; and
- Part 3: Privacy enhancing measures to protect users against mobile phone information-gathering.

## **2. Surveillance and the tension between privacy and security**

In 2013, Wicker (2013) indicated that: "If an information technology collects personal data, governments, law enforcement and commercial interests will exploit the data to its fullest". It is not only the state that wants access to mobile phone data but criminals, tech companies and businesses do too, but they all want it for different purposes. During the 2020 Covid19 pandemic, governments for example turned to mobile phone surveillance in curtailing the spread of the virus. The privacy violation may be justifiable in the interest of public health. However, what happens when Covid19 has been contained? Will the surveillance mechanisms be removed and the data deleted?

Some people may argue that in this era of digital interconnectedness, privacy is nothing more than a myth in a technology-driven society (Veciana-Suarez, 2020). However, such an assumption is not correct (Richards, 2015; Austin, 2015). If privacy was 'dead', then there would not have been much interest in Snowden's revelations (Richards, 2015). Richards (2015) correctly indicates that the collection, use and analysis of many kinds of information is inevitable today. He indicates that privacy relates to the rules that should govern the collection and use of personal information (Richards, 2015).

A mobile phone user, for example, consents to the transfer of information to the service provider to ensure functionality but the user does not give consent to the service provider to sell the data to marketers or to hand it over to law enforcement agencies without a warrant (Wicker, 2013). Likewise a business or a tech company cannot access and/or use personal information without consent in terms of consumer privacy protection legislation.

Privacy protection by means of state and corporate safeguards ensures the security of personal information against unlawful access, collection, use and disclosure. The discussion hereafter will show that privacy and security is currently under threat as a huge amount of information is available and many role players may wish to gain access to this valuable commodity.

## **3. Part 1: Law enforcement access to mobile phone information (state surveillance)**

### **3.1 Introduction**

Electronic evidence gathering forms a major part of criminal investigations either for law enforcement or national security purposes. The purpose of a law enforcement investigation would be to institute a criminal trial and therefore the mobile phone evidence must be gathered in accordance with the correct procedure to ensure admissibility in court.

The issue of privacy and public safety came under the spotlight with the 2013 Snowden revelations which had a negative impact on the ability of law enforcement agencies' to access mobile phone data which is still experienced today.

Snowden, a former National Security Agency (NSA) contractor, leaked classified evidence that the NSA was intercepting and collecting bulk telephone records regarding American telephone calls from U.S. telecommunications providers without warrants (Levinson-Waldman, 2017). Seven years later in September 2020 the U.S. Supreme Court of Appeals for the Ninth Circuit found that the NSA's warrantless telephone dragnet that secretly collected millions of Americans' telephone records violated the Foreign Intelligence Surveillance Act (Sarter, 2019). This ruling does not mean that surveillance, whether state or non-state surveillance, has now ceased to exist but it does confirm the importance of applying surveillance within legislative and constitutional parameters.

Weinstein, Moore and Silverman (2017) made a valid point when they indicated that Snowden's revelation of NSA's bulk collection and surveillance activities impacted negatively on electronic evidence gathering activities by law enforcement agencies. Despite the fact that law enforcement agencies had nothing to do with the indiscriminate collection of data by NSA, a major consequence of the 2013 Snowden leaks was that cell phone and internet providers distanced themselves from law enforcement. Post Snowden, mobile service providers and Internet service providers are of the opinion that law enforcement should obtain their evidence from the user (Weinstein, Moore and Silverman, 2017). It will be illustrated hereafter that law enforcement cannot effectively gather evidence without the co-operation of service providers (Weinstein, Moore and Silverman, 2017). The issue of law enforcement access to encrypted mobile phone communications has been contentious (see par 5.1 hereafter).

### **3.2 Mobile phone location tracking by means of the phone provider**

Mobile phone operators routinely collect Call Detail Records (CDRs) that contain a timestamp and GPS location with a unique identifier for all subscribers. When a user's phone is on, it connects to a cell tower which can determine the location of the user. The only time that it does not have a record is when the phone is switched off. The takeaway from all of this is that when mobile phones are on, they are using data to receive emails, text messages and calls.

Location mobile data may be very useful for crime investigation in the following ways:

- Law enforcement agencies may use mobile communication between the various accused before, during or after the crime along with tower location as a reactive forensic tool to investigate the commission of crime.
- Law enforcement may acquire the suspect's movements without leaving their desk (Pell, 2017). In July 2020, Ghislaine Maxwell, a companion of Jeffery Epstein, was arrested for recruiting and grooming underage girls for sex. She tried to evade arrest by going into hiding. Her location was determined by means of GPS and stingray technology (see par. 3.3 hereafter). The location data established by means of the GPS warrant was insufficient to identify the particular building in which she was hiding. It necessitated the police to apply for a warrant to use a stingray which pinpointed the alleged accused's exact location (Shelton, 2021).
- It may be used to negate a defence. In 2006 a well-known South African musician and playwright was killed in a contract killing which his wife had arranged. She alleged that she was asleep and nowhere near the victim at the time of the killing. The house triggered two cell towers and her mobile phone pinged on the tower in that part of the house where the deceased died which placed her at the crime scene. (Schmitz, Riley and Dryden, 2019).
- It may be used to establish who participated in a crime or were near a crime scene. On 6 January 2021 thousands of pro-Trump supporters crossed the National Mall, overran Capitol Police metal barricades and barged through the halls of the U.S. Capitol building. Law enforcement investigators used mobile phone data to identify individual users whose phones sent out signals from inside the Capitol (Timberg, Harwell and Hsu, 2021).
- Law enforcement may request a provider for a cell tower dump which is a list of mobile devices that were near a certain area at a specific time. In 2014, the Ukrainian government allegedly used this method to determine who was present at an anti-government protest (Brantley, 2014). It may be argued that this type of state surveillance in respect of a peaceful non-violent protest may violate a mobile phone user's constitutional rights, such as the right to free speech and right to movement (see par. 3.3 hereafter).



Despite the advantages of data location tracking, there are some deficiencies. As indicated, the pinging of a mobile phone as well as GPS technology can be used to track a user. However, a breach could occur at several points along the pinging chain. For example, it is possible that someone who has legitimate access to a pinging platform, may forward the login details to others to give them access to it or ineffective oversight from mobile operators could result in location-based service (LBS) abuse. The LBS is used to address crime but it can be abused and therefore it need to be regulated. This was highlighted in September 2020 when a top ranking police official in the South African Anti-Gang Unit, Lieutenant-colonel Kinnear, was murdered in what appears to have been a hired killing. It transpired that the suspect tracked the deceased police official's mobile phone for months and on the day of his murder, the phone pinged more than 2000 times. The abuse of location-based services provided by third party Wireless Application Services Providers (WASPs) went mostly unnoticed in South Africa until this killing (Cruywagen, 2020). He was investigating corruption which allegedly involved police officials. In South Africa tracker companies are authorised to ping SIM cards when a tracker unit stops transmitting (Cruywagen, 2020).

### **3.3 Mobile phone location tracking by means of an IMSI or stingray**

Instant Mobile Subscriber Identity-catchers (IMSI-catchers), also known as stingrays, allow police to track mobile phones and intercept text messages, calls and other data within their radius in real time by imitating cell phone towers (Pell, 2017). IMSI-catchers have been used on the ground and in the air by law enforcement for years but are highly controversial because they do not just collect data from targeted phones; they collect data from any phone in the vicinity of a device. That data can be used to identify people, for example, protesters and track their movements during and after demonstrations, as well as identify others who associate with them. They can also inject spying software onto specific phones or direct the browser of a phone to a website where malware can be loaded onto it.

It is alleged that during the 2020 US Black Lives Matter protests law enforcement agencies used IMSI-catchers (Laperruque, 2020). As indicated, such surveillance may violate protesters' constitutional rights.

## **4. Part 2: Access to consumer' information by companies, businesses and social media messaging apps (non-state/surveillance capitalism)**

### **4.1 Introduction**

Society functions in an age of surveillance capitalism which may be described as a market driven process where the commodity for sale is personal data, and the capture and production of this data relies on mass surveillance of the Internet (Austin, 2017; Holloway, 2019). It will also include the WhatsApp messaging app (see par. 4.2 hereafter).

Organisations have always collected and used data, but organisations have now enhanced surveillance abilities (Austin, 2015). The information-gathering is often carried out by companies that provide users with free online services, such as search engines (i.e. Google) and social media platforms (i.e. Facebook and WhatsApp). In return for providing these free services, these companies collect and scrutinise users' online behaviour (likes, dislikes, searches, social networks, purchases) to produce data that can be further used for commercial purposes. The data used in this process is often collected from the same groups of people who will ultimately be its targets. For instance, Google collects personal online data to target users with advertisements, and Facebook is likely selling users' data to organisations who then may send targeted advertisements to the user (Richards, 2015; Holloway, 2019; also see par. 4.2 hereafter).

Legislation requires that a user consents to the use of personal information. It may be that the user consents to the information-gathering without giving it much consideration. It seems users may be willing to give up their private information in return for perceived benefits such as ease of use, navigation and access to friends and information. However, access, collection, use and disclosure carry with it privacy and security risks. It could be argued that the consent is not an informed consent taking into consideration the possible consequences of such surveillance.

## **4.2 Mobile phone messaging app: WhatsApp**

In early 2021, WhatsApp users received an updated terms of service and privacy policy from Facebook-owned WhatsApp. According to these changes, WhatsApp will share its users' personal information, including phone numbers, IP addresses and contacts with Facebook and its subsidiaries.

In accordance with the new privacy amendments, WhatsApp chats and calls are still end-to-end encrypted, so neither Facebook nor WhatsApp can directly access messages (see par. 5.1 hereafter). The new privacy changes allows profiling of a WhatsApp user where Facebook and its subsidiaries may use this information.

Facebook may be called a free service but in reality it is not (Richards, 2015). Consumers may not pay to use the Facebook service but they cannot use it without giving Facebook the right to collect and use often a vast amount of personal information about them. The business model of Facebook is targeted at advertising. Although WhatsApp may be a free instant messaging app, Facebook could with the new changes obtain and sell the WhatsApp information for advertising purposes. The new changes will enable users to interact directly with businesses for shopping purposes and for businesses to advertise directly to WhatsApp users (Bhengu, 2021).

The policy changes apply everywhere except the European Union (EU) where WhatsApp and United Kingdom has to function in accordance with the General Data Protection Regulation (GDPR) guidelines which outlines strict personal privacy protections (Mehrotra, 2021). If a WhatsApp user does not wish to consent to the use of personal information, then the user has to move to a service such as Signal or Telegram as these messaging apps do not use data linked to users (Mehrotra, 2021). WhatsApp may feel confident that they will not lose a lot of users taking into consideration that in 2020 WhatsApp had 2 billion users (Lin, 2020). It is one of the most popular messaging app today (Lin, 2020).

It is concerning that so much information of WhatsApp users will be shared with Facebook and its subsidiaries as it may have privacy and security implications. WhatsApp users will have to trust Facebook that the data-sharing will be done in a responsible manner. Unfortunately Facebook does not have a good record pertaining to privacy protection. One example of a gross privacy violation is the 2018 Facebook–Cambridge Analytica data scandal in which the personal data of millions of Facebook users were acquired without their consent by the British consulting firm, Cambridge Analytica, predominantly to be used for political advertising (Holloway, 2019). The drawbacks of data-sharing such as the security and privacy risks may outweigh the benefits of WhatsApp usage. Some users may decide to move to Telegram or Signal as they may not want to exchange privacy for a free service.

## **5. Part 3: Privacy enhancing measures to protect users against access of mobile phone information**

### **5.1 Encryption**

There has been conflict between law enforcement agencies and service providers for some time pertaining to law enforcement access to end-to-end encrypted communications. Law enforcement agencies claim that encryption and other privacy technologies cause their surveillance to “go dark” (Gray and Henderson, 2017). But industry leaders have warned that any system requiring a “backdoor” to encryption would undermine the privacy protections altogether.

The tech industry has fought to maintain the integrity of encryption which prevents even the companies that make the devices or platforms from being able to access their contents. Encrypted services only let the sender and recipient see messages. Law enforcement officials have insisted that they must have some way to access encrypted platforms and devices when investigating crimes.

In 2020, a group of U.S. Republican senators introduced a bill, the Lawful Access to Encrypted Data, that would weaken the lawful use of encryption in order for law enforcement officials to gain access to devices and communication services with a warrant (Marks, 2020). The European Data Protection Supervisor noted in 2020 that the confidentiality of communications is a cornerstone of the fundamental rights to respect for private and family life (Nielsen, 2020). However, EU law enforcement agencies wish to gain access to encrypted information to investigate serious crimes such as extremism and child abuse. The European Commission indicated in 2020

that it may introduce new EU-rules on end-to-end encryption, possibly allowing police to crack into platforms like WhatsApp or Signal (Nielsen, 2020).

User privacy and security must be weighed against the law enforcement need to investigate serious crimes and in some instances, public safety will outweigh privacy rights (Watney, 2020).

## **5.2 Apple App Store privacy requirements**

Apple has indicated that from 2021 it will require apps to reveal how user data may be collected (Whitney, 2020). Any new or updated app must include a privacy label, otherwise it will not be available on the App Store. This requirement applies not just to third-party apps but to Apple's own programs, such as Apple Music, Apple TV and Apple Wallet, though built-in apps are not included. The goal is to address privacy concerns and questions among users, especially as app developers have not always explained clearly and precisely which data they collect and how they use it.

The App privacy labels will provide users with greater transparency as they will know how their data is used, the obligations providers have to protect that data and the controls individuals have to monitor, disable, and delete their data.

The biggest complaint so far has come from Facebook, which pointed to several pitfalls that it sees in Apple's new privacy process. Facebook indicated that the format of Apple's new labels is too broad and ignores how data is used in context (Whitney, 2020). Facebook purchased, at the end of 2020, a series of full-page newspaper advertisements directed at Apple — a direct attack on the iPhone maker's new ad tracking policies. Facebook's new ads reiterate CEO Mark Zuckerberg's claims that Apple's tracking policies will hurt small businesses most (Wille, 2020). To some extent, Facebook is not entirely wrong in its assertions. Apple's updated tracking policies will affect small businesses. Users will be given the option of refusing ad tracking on an app-by-app basis — a change that will affect any business that utilizes personalized ads. However, it is not only small businesses but also Facebook that will lose profits because of the change. Facebook's ad empire is worth upwards of \$70 billion at last count, and many of those ads rely on user-tracking (Wille, 2020). Facebook makes no mention of its own interest in this cause and in the advertisement it only refers to small businesses. It appears that Facebook is pushing back at Apple in a public way because its profit margins will be hurt by the new ad tracking policies. Restrictions on what information businesses may collect or how they may use the information cut into their profits (Richards, 2015).

It is interesting that in 2021 WhatsApp proposed an amendment to their privacy policy to provide for data-sharing with Facebook (see par. 4.2). The amendments to WhatsApp may be in response to Apple Store's new privacy labels and transparency policy. Access to WhatsApp users' information will ensure that Apple Store's privacy changes do not impact too negatively on Facebook's business model of targeted advertisements.

## **6. Conclusion**

### **6.1 Law enforcement (state) surveillance**

Law enforcement mobile phone information-gathering results in a trade-off between privacy and public safety. As Weinstein, Moore and Silverman (2017) justifiably indicate "We cannot have a perfect version of either one without some compromise of the other."

Privacy advocates and providers are correct that user information should be protected, that encryption is a valuable service that customers are increasingly demanding and that there are fewer risks if encryption does not include a mechanism for government access. Law enforcement on the other hand is correct that there are public safety consequences when it cannot access information. One can support both sides at the same time (Weinstein, Moore and Silverman, 2017). If one supports strong encryption as a privacy enhancing tool, it does not mean that one does not support public safety. Law enforcement understands the importance of privacy protection and in this regard, evidence need to be gathered in a lawful manner by means of a warrant otherwise it may be inadmissible in a criminal court and allow criminals to evade justice.

## 6.2 Non-state/corporate) surveillance

Austin (2015) accentuates the impact of surveillance capitalism on privacy and security. As indicated, users must give consent for information-gathering but is it informed consent? As we move towards 5G technology and IoTs against the background of the Fourth Industrial Revolution, access to big data is a serious concern. It is important to establish who has access to data and also for what purpose and when. Here mobile phone information security is essential as it is possible for cybercriminals, for example, to get hold of personal information.

## 6.3 Surveillance in general

Personal information has considerable value to both law enforcement agencies and companies. The collection, use and disclosure of personal information is inevitable (Richards, 2015). State and non-state surveillance may come down to drawing a line between 2 values, privacy and security which are both equally important. It is important to reflect on the circumstances in which these values may be limited and the justifiability of such limitation (Weinstein, Moore and Silverman, 2017). Gray and Henderson (2017) indicate that there will never be a “neat and tidy solution” pertaining to surveillance. A solution needs to balance competing interests which shifts and changes with society and technology (Gray and Henderson, 2017).

Legal and social rules are essential in respect of regulating how information may be obtained and used (Richards, 2015). Austin (2015) indicates that the expansion of surveillance may not be prevented by means of privacy and data protection regimes. Surveillance is part of the technological digital era but constant and critical oversight of surveillance may negate abuse and/or exploitation of personal information.

## References

- Austin, L.M. (2015) “Enough about me: Why privacy is about power, not consent (or harm)”, *A world without Privacy*, Cambridge University Press, United Kingdom, pp. 131, 144, 148 – 149, 218-219, 229 – 230, 233.
- Bankmycell. (2021) “How many smartphones are in the world?”, [online], <https://www.bankmycell.com/blog/how-many-phones-are-in-the-world>.
- Bhengu, S. (2021) “Goodbye WhatsApp? Here's why your friends are flocking to Signal & Telegram”, [online], <https://www.timeslive.co.za/news/sci-tech/2021-01-11-goodbye-whatsapp-heres-why-your-friends-are-flocking-to-signal-telegram/>.
- Brantley, A. (2014) “You were identified as a participant in a mass disturbance”, [online], <https://www.nditech.org/you-were-identified-participant-mass-disturbance>.
- Cruywagen, V. (2020) “Former rugby player arrested in Kinnear investigation set to appear in Cape Town court”, [online], <https://www.dailymaverick.co.za/article/2020-09-24-former-rugby-player-arrested-in-kinnear-investigation-set-to-appear-in-cape-town-court/>.
- Gray, D. and Henderson, S. (2017) “Introduction”, *The Cambridge Handbook of Surveillance Law*, Cambridge University Press, Cambridge, p 1 – 3.
- Holloway, D (2019) “Explainer what is surveillance capitalism and how does it shape our economy”, [online], <https://theconversation.com/explainer-what-is-surveillance-capitalism-and-how-does-it-shape-our-economy-119158>.
- Laperruque, J. (2020) “How to Respond to Risk of Surveillance While Protesting”, [online], <https://www.pogo.org/analysis/2020/10/how-to-respond-to-risk-of-surveillance-while-protesting/>.
- Levinson-Waldman, R. (2017) “NSA Surveillance in the War on Terror”, *The Cambridge Handbook of Surveillance Law*, Cambridge University Press, Cambridge, p 7.
- Lin, Y. (2020) “10 WhatsApp statistics every marketer should know in 2021”, [online], <https://www.oberlo.co.za/blog/whatsapp-statistics>.
- Marks, J. (2020) “Cybersecurity pros are uniting in battle to save encryption”, [online], <https://www.iol.co.za/technology/software-and-internet/cybersecurity-pros-are-uniting-in-a-battle-to-save-encryption-50610459>.
- Mehrotra, P. (2021) “WhatsApp updates its terms and conditions to mandate data-sharing with Facebook”, [online], <https://www.xda-developers.com/whatsapp-updates-terms-privacy-policy-mandate-data-sharing-facebook/>.
- Mzemkandaba, S. (2020) “SA’s smartphone penetration surpasses 90%”, [online], <https://www.itweb.co.za/content/xA9PO7NZRad7o4J8>.
- Nielsen, N. (2020) “EU Commission mulls police access to encrypted apps”, [online], <https://euobserver.com/justice/150334>.
- Pell, S.K. (2017) “Location tracking”, *The Cambridge Handbook of Surveillance Law*, Cambridge University Press, Cambridge, pp 44-45, 69.
- Richards, N.M. (2015) “Four Privacy Myths”, *A world without Privacy*, Cambridge University Press, United Kingdom, pp. 22, 35, 43, 50, 72, 74, 81.

## **Murdoch Watney**

- Sarter, R. (2019) "U.S. Court – Mass surveillance program exposed by Snowden was illegal", [online] <https://uk.reuters.com/article/uk-usa-nsa-spying/u-s-court-mass-surveillance-program-exposed-by-snowden-was-illegal-idUKKBN25T3D2?il=0>.
- Schmitz, P.M.U., Riley, S. and Dryden, J. (2009) "The use of mapping time and space as a forensic tool in a murder case in South Africa", [online], [https://icaci.org/files/documents/ICC\\_proceedings/ICC2009/html/refer/20\\_5.pdf](https://icaci.org/files/documents/ICC_proceedings/ICC2009/html/refer/20_5.pdf).
- Shelton, T. (2021) "FBI located Ghislaine Maxwell by tracking her mobile phone, court documents show", [online], <https://www.abc.net.au/news/2021-01-06/fbi-located-ghislaine-maxwell-by-tracking-her-mobile-phone/13035274>.
- Timberg, C., Harwell, D. and Hsu, S.S. (2021) "Video, cellphone and facial recognition may lead police to Capitol rioters", [online], <https://www.adn.com/nation-world/2021/01/08/video-cellphone-and-facial-recognition-data-may-lead-police-to-capitol-rioters/>.
- Veciana-Suarez, A. (2020) "Privacy is nothing more than a myth in the digital age", [online], <https://www.miamiherald.com/living/liv-columns-blogs/ana-veciana-suarez/article239094908.html>.
- Watney, M.M. (2020) "Law Enforcement Access to End-to-End Encrypted Social Media Communications", 7<sup>th</sup> *European Conference on Social Media*, Academic Conference and International Conference Limited, Reading, UK, pp. 332 -329.
- Weinstein, J. M., Moore R.J. and Silverman, N.P. (2017) "Balancing Privacy and Public Safety in the Post-Snowden Era", *The Cambridge Handbook of Surveillance Law*, Cambridge University Press, Cambridge, pp 227, 238 – 239, 241, 246 -247.
- Whitney, L. (2020) "How Apple's new App Store privacy requirements may affect users and app developers", [online], <https://www.techrepublic.com/article/how-apples-new-app-store-privacy-requirements-may-affect-users-and-app-developers/>.
- Wicker, S.B. (2013) *Cellular convergence and the death of privacy*, Oxford University Press, New York, US, pp, 4, 13, 15, 19, 63, 170, 173, 174.
- Wille, M. (2020) "Facebook bought full-page print ads to hide its greed behind small businesses", [online], <https://www.inputmag.com/culture/facebook-bought-full-page-ads-to-hide-behind-small-businesses>.



# **PhD Research Papers**





# The Impact of GDPR Infringement Fines on the Market Value of Firms

Adrian Ford<sup>1</sup>, Ameer Al-Nemrat<sup>1</sup>, Seyed Ali Ghorashi<sup>1</sup> and Julia Davidson<sup>2</sup>

<sup>1</sup>School of Architecture, Computing and Engineering, University of East London, UK

<sup>2</sup>Royal Docks School of Business and Law, University of East London, UK

[a.ford1701@uel.ac.uk](mailto:a.ford1701@uel.ac.uk)

DOI: 10.34190/EWS.21.088

**Abstract:** Previous studies have shown (varying degrees of) evidence of a negative impact of data breach announcements on the share price of publicly listed companies. Following on from this research, further studies have been carried out in assessing the economic impact of the introduction of legislation in this area to encourage firms to invest in cyber security and protect the privacy of data subjects. Existing research has been predominantly US centric. This paper looks at the impact of the General Data Protection Regulation (GDPR) infringement fine announcements on the market value of mostly European publicly listed companies with a view to reinforcing the importance of data privacy compliance, thereby informing cyber security investment strategies for organisations. Using event study techniques, a dataset of 25 GDPR fine announcement events was analysed, and statistically significant cumulative abnormal returns (CAR) of around -1% on average up to three days after the event were identified. In almost all cases, this negative economic impact on market value far outweighed the monetary value of the fine itself, and relatively minor fines could result in major market valuation losses for companies, even those having large market capitalisations. A further dataset of four announcements where sizeable GDPR fines were subsequently appealed was also analysed and although positive returns for successful appeals were observed (and the reverse), they could not be shown to be statistically significant - perhaps due, at least in part, to COVID-19 related market volatility at that time. This research would be of benefit to business management, practitioners of cyber security, investors and shareholders as well as researchers in cyber security or related fields (pointers to future research are given). Data protection authorities may also find this work of interest.

**Keywords:** cyber security, data privacy breaches, market value, economic impact, GDPR, event study

---

## 1. Introduction

The European Union Agency for Cybersecurity (ENISA, 2020) reported a “54% increase in the total number of [data] breaches by midyear 2019 compared with 2018”. Regarding the introduction of the General Data Protection Regulation (GDPR) in May 2018, ENISA also remark that “55% of the responders to a Eurobarometer survey responded that they are concern[sic] about their data being accessed by criminals and fraudsters”. Clearly there is major concern out there in the field of data privacy. The primary objective of the GDPR is to protect “fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data” (Data Protection Act 2018). The requirement, therein, to notify data breaches to the relevant supervisory authority within 72 hours of becoming aware (where feasible), could reasonably be expected to increase visibility of non-compliance. For example, in the UK, before the introduction of the GDPR as the Data Protection Act (DPA), 2018, the preceding DPA (1998), according to the Information Commissioner’s Office (ICO)<sup>1</sup>, stated “although there is no legal obligation on data controllers to report breaches of security which result in loss, release or corruption of personal data, the Information Commissioner believes serious breaches should be brought to the attention of his Office.” Prior to 2010, the ICO were limited to serving enforcement notices for contraventions of the DPA (1998), however in April 2010 the ICO was granted the power to issue fines of up to £500,000 on its own authority. For example, Sony Computer Entertainment Europe were fined £250,000 in January 2013 for a “serious breach” when their PlayStation Network was hacked (BBC 2013) and in 2016, TalkTalk were fined £400,000 for leaking personal data of almost 157,000 customers due to poor website security (BBC 2016). Serious infringements under the GDPR, those violating the fundamental principles of the right to privacy and the right to be forgotten, could result in a fine of up to €20 million or 4% of the firm’s worldwide annual revenue from the preceding financial year (whichever amount is higher), a clear deterrent against carelessness concerning data privacy and security. Indeed, total fines issued by data protection authorities since the introduction of the GDPR currently stand at over €275m (CMS Legal 2021).

This research is concerned with the impact the announcement of such GDPR fines has on the market value of publicly listed companies. Spanos and Angelis (2016) report that data breach announcements are associated with a negative impact on market value. Could it be that, since the introduction of the GDPR, a firm’s share price

---

<sup>1</sup> The supervisory (data protection) authority of the UK (<https://ico.org.uk>)

may suffer a ‘double whammy’ of both initial breach notification and subsequent punitive action? This paper aims to assess the economic impact of the introduction of the GDPR on publicly listed companies through the application of fiscal penalties levied by its supervisory authorities on those firms which have suffered a data privacy breach. By gaining a greater understanding in this area it is hoped to encourage firms to invest more in cyber security measures to prevent such occurrences. To achieve this objective, the following research questions were considered:

- Is there any impact on company market value of a publicly announced GDPR fine?
- Do data analyses reveal any obvious patterns/correlations?
- What is the impact of any fine successfully appealed and subsequently overturned or reduced?
- How can the results inform cyber security investment strategies?
- Can any conclusions be drawn about the introduction of the GDPR itself?

This research will highlight the importance of data privacy and protection to business management and thus the need to invest in and improve their organisation’s cyber security posture<sup>2</sup> thereby reducing the risk of data privacy breaches. Such insight would also assist practitioners of information security with business case justifications. This research would be of benefit to business management, practitioners of cyber security, investors and shareholders as well as researchers in cyber security or related fields. It could also be of value to data protection authorities to increase their understanding of the impact and enforcement of legislation on the economy. Another benefit of this study would be the European focus thereby beginning to offset the strong US bias of the existing literature in this area.

## **2. Related work**

A systematic literature review concerning the impact of data breach events on the stock market carried out by Spanos and Angelis (2016) reports that, although research in this area was “*quite limited*”, the majority of studies (76%) found a statistically significant negative impact. For example, Lin et al. (2020) report a loss of 1.44% on average over a 5-day window. Andoh-Baidoo, Amoako-Gyampah and Osei-Bryson (2010) report -3.18% abnormal returns over a 3-day period. Cavusoglu, Mishra and Raghunathan (2004) cite -2.1% on average within two days after the announcement. Goel and Shawky (2009) quote -1% in the days surrounding the event. These studies also note some correlations between these negative returns and, for example, industry sector. Tweneboah-Kodua, Atsu and Buchanan (2018), warn that “*studying the cumulative effects of cyberattacks on prices of listed firms without grouping them into the various sectors may be non-informative*”. They noted that financial services firms reacted more rapidly and more significantly than those in the technology sector. It was also observed by Campbell et al. (2003) that those breaches involving unauthorised access to confidential data were more likely to result in significant negative market reaction, which one would reasonably expect to apply across the board for this study. Such observations would support the idea of governments introducing legislation to not only counter this negative economic impact but also to help protect data subjects who are effectively innocent victims of such breaches of confidentiality. Indeed, the right to privacy is a component of the European Convention on Human Rights (1950) and the EU has sought to protect this right through legislation ever since with, firstly, the introduction of the European Data Protection Directive (1995) then the Privacy and Electronic Communications Directive (2002) and, in response to ever-evolving technology and increases in data transfers, the GDPR in 2018 along with the (delayed) ePrivacy Regulation due to repeal the 2002 Directive (European Commission 2021).

This relatively recent introduction of the GDPR naturally limits the availability of research on its impact, so it is necessary to look elsewhere. The introduction of data breach notification laws in the US was found to reduce identity theft by over 6% on average (Romanosky, Telang & Acquisti 2011). Clearly if data subjects are rapidly made aware their personal data has been compromised, and which data, they should be better positioned to take preventative action. There are already, however, some criticisms of the effectiveness of the GPDR in this area as notification to data subjects is only required in certain “*high risk*” cases and where it would not place too onerous a burden on the reporting organisation (Nieuwesteeg & Faure 2018). Data breach notification laws have been widely adopted in the US, albeit not centrally – federal law in this area only covers certain specific sectors. Nevertheless, 47 jurisdictions have implemented their own notification legislation. In fact, the US could be considered an early adopter. In contrast the EU GDPR model is central and adopted by member states and

---

<sup>2</sup> Cyber security posture includes not only governance and technical solutions but also training and awareness.

includes the notification requirement within the data protection law itself unlike, for example, Australia (Daly 2018) where a separate law was introduced in early 2017. Goel and Shawky (2014) carried out a US based study examining the impact of data breach announcements on share price and found a significant reduction in negative returns after the enactment of both federal and state laws. Murciano-Goroff (2019) looked at Californian company investment in web server security following the introduction of state data breach notification law yet only noted a modest effect with server software being, at most, 2.8% newer. Indeed, Richardson, Smith and Watson (2019) argue that “companies are unlikely to change their investment patterns unless the cost of breaches increases dramatically or regulatory bodies enforce change” underpinning the need for an understanding of the impact and effectiveness of the GDPR on cyber security investment – an area which this research aims to inform as well as bringing an EU specific perspective to offset the strong US bias of previous studies.

### 3. Methodology

The high-level approach to this research was to download a list of publicly announced GDPR infringement fines from the Enforcement Tracker (CMS Legal 2021), filter this dataset for those cases involving publicly listed companies and analyse the impact of these announcements on share price using event study techniques.

#### 3.1 Event studies

Event studies have been widely used to assess the impact of specific events on the share price of firms and thereby their market value and are described in detail in, for example, MacKinlay (1997). A key assumption of this methodology is the ability of the market to reflect all available information as per the efficient market hypothesis (e.g. Fama 1970). By observing share price movements in reaction to information regarding a specific event, such as a data breach announcement over a short time period (the event window) it is possible to deduce how the market reacted to that specific event, given there are no other confounding events during that time-period.

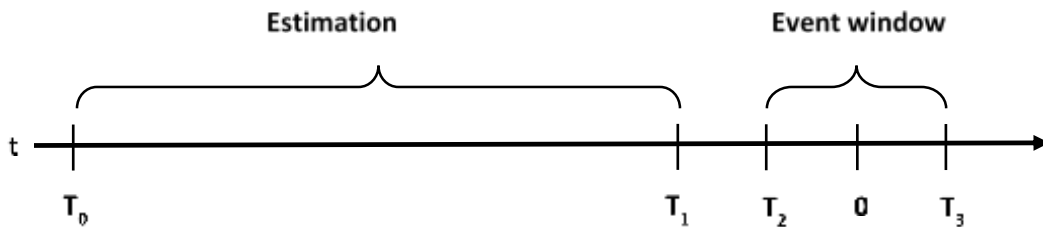


Figure 1: Event study timeline

A common approach used in similar (data breach type) event studies is the market model (e.g. Cavusoglu et al. 2004; Andoh-Baidoo et al. 2010; Hinz et al. 2015; Schatz & Bashroush 2016; Castillo & Falzon 2018; Tweneboah-Kodua et al. 2018; Jeong et al. 2019) which uses an estimation window prior to the (shorter) event window (see Figure 1) to predict movement of the firm’s stock based on a regression analysis. Returns are assumed to follow a single factor model (1) where the return of firm  $i$  on day  $t$  ( $R_{i,t}$ ) is dependent on the corresponding daily return of the reference market ( $R_{m,t}$ ) and the extent of the security’s responsiveness ( $\beta_i$ ) offset by its abnormal return ( $\alpha_i$ ). The error term  $\varepsilon_{i,t}$  is expected to be zero with finite variance. Abnormal returns are calculated for the event window (2) and reported as a cumulative abnormal return (CAR) over the whole event window (3). For cross-sectional analyses a cumulative average abnormal return (CAAR) was calculated for  $N$  events as shown in equation (4).

$$R_{i,t} = \alpha_i + \beta_i \cdot R_{m,t} + \varepsilon_{i,t} \tag{1}$$

$$AR_{i,t} = R_{i,t} - (\alpha_i + \beta_i R_{m,t}) \tag{2}$$

$$CAR_i = \sum_{t=T_2}^{T_3} AR_{i,t} \tag{3}$$

$$CAAR = \frac{1}{N} \sum_{i=1}^N CAR_i \tag{4}$$

### 3.2 Data collection

The base dataset used to identify fine announcements was from the GDPR Enforcement Tracker. Although not professing to be an exhaustive list, when the data were downloaded in May 2020 this resulted in 277 records. Manually filtering these records for those involving publicly listed companies (or a subsidiary of a publicly listed company<sup>3</sup>) resulted in 71 rows. Some announcement dates were found to be missing and filled in from press reports and official data protection authority publications where applicable. It was necessary to exclude certain records due to a missing date such as Facebook (Germany) and Unicredit (Czech Republic/Slovakia). Events on the same day were consolidated into one e.g. Eni Gas e Luca, EDP Spain. Entries which had potentially overlapping event windows were also filtered e.g. Vodafone (2 events). Share price and market index data were extracted from Yahoo!Finance (2019) along with firm demographics such as annual revenue, market capitalisation and industry sector. Information was not available for all the events on Yahoo!Finance e.g. Louis Group (Cyprus), Xfera (now privately owned) and Avon Cosmetics (event was pre-public), thus these events had to be filtered out also, leaving 48 records. The most appropriate market index was chosen as a reference in each case (Kannan, Rees and Sridhar (2007) highlighted the importance of the market reference), ideally one which included the candidate company itself but adjusted, if needed, due to lack of availability of data in Yahoo!Finance. Some firms had multiple listings in which case the primary listing and associated index were used. The date range was limited, naturally, from the earliest fine since the introduction of the GDPR in 2018 (actually, January 2019) until the date of download but it was decided to cap the data at 31/12/2019 in order to avoid market uncertainties due to COVID-19, that being a long-term confounding event in itself. This date capping reduced the dataset from 48 to 25 events for analysis.

### 3.3 Data analysis

To facilitate the analyses, R (R Core Team 2018)<sup>4</sup> scripts were developed to pull share price and index data directly from Yahoo!Finance for each data record and then event studies run using an R package (Schimmer, Levchenko & Müller 2014)<sup>5</sup> using the market model as described above. Non-trading event days were defaulted to the next available trading day. An estimation window of 120 days was chosen consistent with e.g. Goel and Shawky (2009), Andoh-Baidoo et al. (2010), Schatz and Bashroush (2016), Richardson et al. (2019). In all cases the estimation window ended one trading day before the event window. Tweneboah-Kodua et al. (2018) recommend avoiding overlap of the estimation and event windows in this way to avoid “parameter contamination”. Although the event window should be broad enough to contain any uncertainty in the date of the event, the longer the window the less likely it is to detect abnormal returns (Dyckman, Philbrick, & Stephan 1984). Previous studies have shown market reaction before the event date due to information leakage. For example, using event study techniques, Lin et al. (2020) show significant evidence of opportunistic pre-official announcement insider trading related to data breaches. For this study, a range of event windows were initially chosen starting from up to two days before the event and varying in length from 2 up to 20 trading days to give visibility of these effects and others such as sector specific effects reported by e.g. Tweneboah-Kodua et al. (2018) who observed more rapid response from the financial services sector, for instance.

### 3.4 Hypothesis development

For event studies, the null hypothesis maintains that there are no abnormal returns within the event window. The standard deviation of abnormal returns during the event window is described by equation (5) where  $M_i$  refers to the number of non-missing returns. The t-value for the CAR over the event window was then calculated according to equation (6).

$$S_{AR_i} = \sqrt{\frac{1}{M_i - 2} \sum_{t=T_0}^{T_1} (AR_{i,t})^2} \tag{5}$$

$$t_{CAR} = \frac{CAR_i}{\sqrt{(T_3 - T_2 + 1)S_{AR_i}^2}} \tag{6}$$

<sup>3</sup> Ultimate parent companies were identified from Dun & Bradstreet (<https://www.dnb.com>)

<sup>4</sup> R version 4.0.3 (2020-10-10)

<sup>5</sup> EventStudy package version 0.36.900 (API version 0.374-alpha)

For cross-sectional analyses the t-statistic ( $t_{CAAR}$ ) was calculated based on the CAAR as shown in (8) with  $S_{CAAR}$  being the standard deviation of the CARs for each firm  $i$  across the sample of size  $N$  (7).

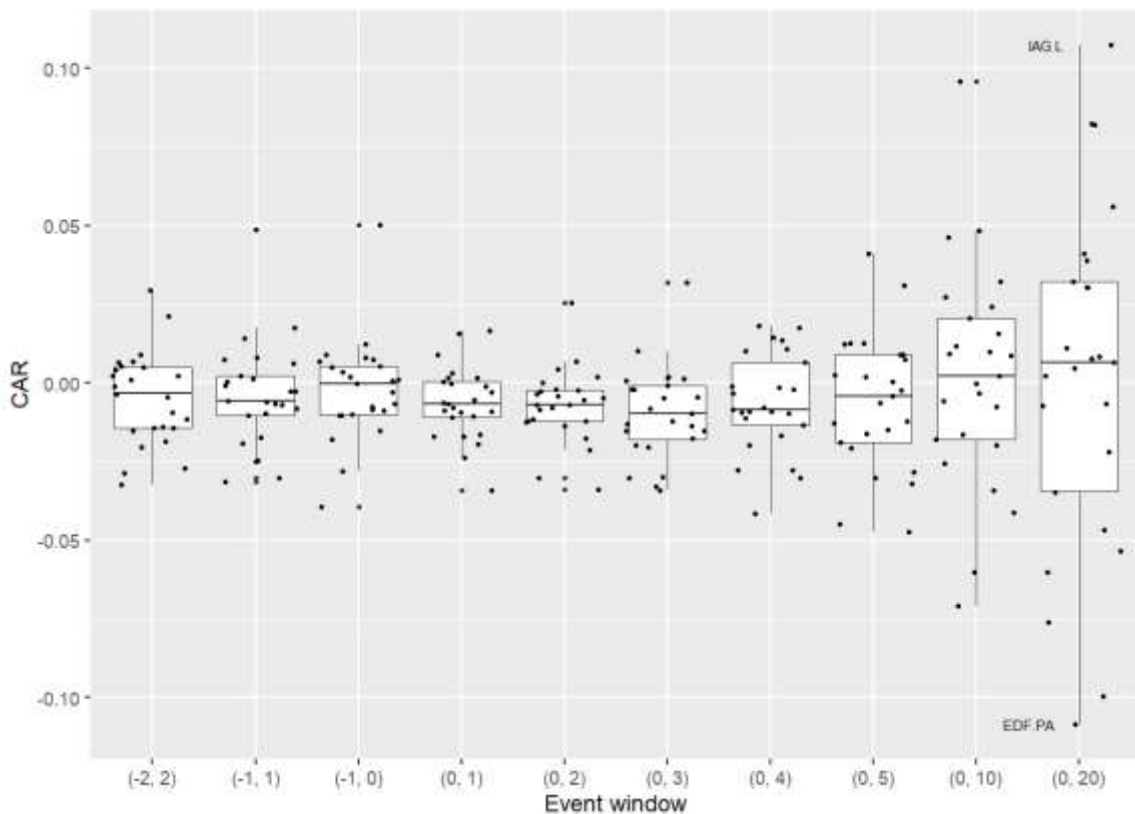
$$S_{CAAR} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (CAR_i - CAAR)^2} \tag{7}$$

$$t_{CAAR} = \sqrt{N} \frac{CAAR}{S_{CAAR}} \tag{8}$$

This approach to significance testing is consistent with e.g. Castillo and Falzon (2018), Deane et al. (2019) and Jeong et al. (2019). Indeed, Deane et al. (2019: 115) state that “the t test is considered to be the best framework for analyzing statistical significance in most event study frameworks and to be relatively robust”.

#### 4. Results and discussion

Event studies were carried out as described above for 10 event windows of varying length across all 25 GDPR fine events. A visualisation of the overall results is shown in **Figure 2**. It appears at first glance that the most negative impact is seen around the 4-day event window (0, 3) with the market value gradually recovering over longer windows and beginning to see positive recovery 10 days after the event. After 20 days, for IAGL (Vueling) and EDF (Madrileña Red de Gas) the abnormal returns had grown to over 10% either way yet the median CAR remained much closer to zero.



**Figure 2:** Comparison of event windows

A CAAR was calculated for multiple firms across each window and is shown in **Table 1**. Here the 3 and 4-day event windows (0, 2), (0, 3) show the most negative abnormal returns and are statistically significant at the 1% level. It is interesting to note that the null hypothesis cannot be rejected for the three earlier event windows involving pre-event days thereby indicating no information leakage prior to the fine announcements and consistent with the lack of uncertainty in the event dates for this exercise. As above, there is also lack of statistical significance for the longer windows indicative of a tendency of market recovery towards zero abnormal returns over time as reported by Dyckman et al. (1984). The event window (0, 3) showed the most

negative (almost 1%) CAAR, consistent with the findings of Goel and Shawky (2009). Within this window, 19 of the 25 events (76%) had abnormal returns of less than zero, therefore this window was chosen as the basis for further analyses. Usage of this event window (0, 3) has been previously reported in studies of this type e.g. Hinz et al. (2015), Rosati et al. (2019).

**Table 1:** CAAR by event window

Event Window	N	CAAR	t <sub>CAAR</sub>		% Negative CAR
(-2, 2)	25	-0.0049	-1.6188		56
(-1, 1)	25	-0.0041	-1.2112		64
(-1, 0)	25	-0.0022	-0.6746		52
(0, 1)	25	-0.0064	-2.7453	**	72
(0, 2)	25	-0.0072	-3.0748	***	80
(0, 3)	25	-0.0096	-3.2341	***	76
(0, 4)	25	-0.0064	-2.0190	*	72
(0, 5)	25	-0.0061	-1.4128		56
(0, 10)	25	0.0020	0.2795		48
(0, 20)	25	0.0011	0.0968		40
	<b>250</b>	<b>-0.0044</b>			<b>62</b>
*,**,*** Represent statistical significance at the 10%, 5% and 1% levels respectively.					

An analysis by ultimate parent company of CAAR is shown in **Table 2**. It can be seen that four firms suffered more than one fine under GDPR, but no more than two during the date range of this study. The firm suffering the most negative abnormal return is listed first and the most positive last. The overall average fine levied was found to be almost €17m and it appears that the supervisory authorities have been relatively lenient so far with the average penalty sitting at around 0.15% of previous year's annual revenue (the greatest being just over 1%) and nowhere near the possible maximum of 4% for more serious GDPR infringements<sup>6</sup>. That said, the average loss in market capitalisation based on the CAAR was estimated to be of the order of nearly 29,000 times that at €1.2bn. Clearly this figure is heavily skewed by the €19bn loss Alphabet Inc. experienced following their €50m fine. It seems that a huge market value is little protection against abnormal returns with the smallest company in the sample, Österreichische Post, having a slightly positive return. Also noteworthy was the seemingly innocuous €2k fine for BNP Paribas precipitating a market value fall of nearly €1bn. It was also noted that there was only one case (Österreichische Post) out of all 25 where the ratio of change in market capitalisation to fine was less than one, so firms need to recognise that the overall financial impact of a GDPR penalty is likely to be much greater than the value of the actual fine itself.

**Table 2:** Analysis by ultimate parent company

Ultimate Parent Company	N	CAAR	Average Revenue † € 000,000	Average Fine € 000	Fine as % of Revenue	Market Capitalisation ‡ € 000,000	Δ Market Capitalisation € 000	Δ MC to Fine Ratio
United Internet	1	-0.0342	5,131	9550	0.1861	7,104	242,957	25
Endesa SA	1	-0.0300	19,555	60	0.0003	22,634	679,020	11,317
Iberdrola	2	-0.0253	35,076	42	0.0001	63,221	1,602,652	38,618
UniCredit	1	-0.0204	20,674	130	0.0006	18,639	380,236	2,925
Delivery Hero	1	-0.0198	665	195	0.0294	23,691	469,082	2,401
Alphabet Inc	1	-0.0153	120,380	50000	0.0415	1,245,280	19,052,788	381
BNP Paribas	1	-0.0152	52,030	2	0.0000	61,513	934,998	467,499
International Airlines	2	-0.0148	24,406	102315	0.4192	10,354	153,246	1

<sup>6</sup> Note that percentages were calculated based on ultimate parent revenues and not necessarily that of the infringing legal entity.

Ultimate Parent Company	N	CAAR	Average Revenue † € 000,000	Average Fine € 000	Fine as % of Revenue	Market Capitalisation ‡ € 000,000	Δ Market Capitalisation € 000	Δ MC to Fine Ratio
Vodafone	1	- 0.0130	43,666	60	0.0001	40,960	532,482	8,875
Eni SpA	1	- 0.0123	75,822	11500	0.0152	33,157	407,831	35
Deutsche Telekom	2	- 0.0110	75,351	21	0.0000	70,219	768,898	36,614
Marriott	1	- 0.0097	18,507	110390	0.5965	41,340	400,995	4
Enel SpA	1	- 0.0049	74,221	6	0.0000	82,095	402,266	67,044
ING Group	1	- 0.0046	18,304	80	0.0004	34,953	160,784	2,010
OTP Bank	1	- 0.0019	2,955	511	0.0173	10,979	20,861	41
Direct Line Insurance	1	- 0.0007	3,937	5	0.0001	4,954	3,468	694
Électricité de France	1	0.0014	68,976	12	0.0000	31,142	43,599	3,633
Engie SA	1	0.0016	60,596	60	0.0001	30,778	49,245	821
Österreichische Post	1	0.0019	1,958	18000	0.9191	2,320	4,408	0
Telefónica	2	0.0042	48,693	39	0.0001	20,019	84,080	2,156
Deutsche Wohnen	1	0.0320	1,438	14500	1.0086	13,665	437,280	30
	<b>25</b>	<b>-0.0096</b>	<b>38,235</b>	<b>16796</b>	<b>0.1462</b>	<b>81,313</b>	<b>1,177,602</b>	<b>28,901</b>

† Revenue of fiscal year prior to the event (consistent with GDPR penalties). Currencies converted based on rate at time of event.

‡ Current market capitalisation (Feb-21). Currencies converted based on rate at 31/12/2019.

Noting that of the top four negative CAAR events in **Table 2**, three of them are related to electricity companies it would certainly be interesting to look at industry sector analysis as recommended by e.g. Tweneboah-Kodua et al. (2018). A breakdown by sector is shown in **Table 3**. Here it can be seen that the most reactive industry sector was *Consumer Cyclical* (-1.5%), however, only *Utilities*, *Communication Services* and *Financial Services* showed statistical significance of non-zero (negative) abnormal returns albeit only at the 10% level.

**Table 3:** CAAR by industry sector

Industry Sector	N	CAAR	t <sub>CAAR</sub>	% Negative CAR
Consumer Cyclical	2	-0.0148	-2.9208	100
Utilities	6	-0.0138	-2.1852	67
Energy	1	-0.0123		100
Communication Services	7	-0.0109	-2.1098	86
Industrials	3	-0.0092	-0.8761	33
Financial Services	5	-0.0086	-2.1881	100
Real Estate	1	0.0320	-2.0190	0
	<b>25</b>	<b>-0.0096</b>		<b>76</b>

\* Represents statistical significance at the 10% level.

During the data collection exercise, it was noted that some of the larger GDPR fines had been appealed and the results of the appeals formally announced. This enabled an additional data set to be built (**Table 4**) and analysed in the same way as the initial announcements.

**Table 4:** Summary of GDPR fine appeals

Ultimate Parent	Date	Original fine	Result of appeal
Alphabet Inc	12/06/2020	€50m	Rejected
International Airlines	16/10/2020	£190m	Reduced to £20m
Marriott	30/10/2020	£99.2m	Reduced to £18.4m
United Internet	12/11/2020	€9.55m	Reduced to €900k

The expected outcome of these appeal announcements would be negative market price impact for the unsuccessful appeal by Alphabet Inc and positive for the other three examples where the fines were massively reduced. The results are shown in **Table 5**. It appears there is indeed, a negative trend for Alphabet beginning on the announcement day itself and not disappearing until 20 days after the event. International Airlines has a strongly increasing positive return after the event whereas, although positive, United Internet remains fairly constant. Marriott however, experienced some negative market sentiment after the event. One has to be mindful of market conditions and volatility due to the COVID-19 pandemic and its effect on (especially the hospitality) industry here. That was the reason the original data set was capped at 31/12/2019 and, in analysing these more recent events, the results were not found to be statistically significant thus the null hypothesis of zero abnormal returns still stands.

**Table 5:** CAR by event window of fines appealed

Event Window	N	Alphabet Inc		International Airlines		Marriott		United Internet	
		CAR	t <sub>CAR</sub>	CAR	t <sub>CAR</sub>	CAR	t <sub>CAR</sub>	CAR	t <sub>CAR</sub>
(-2, 2)	1	0.0164	0.5686	0.1459	1.1842	0.0455	0.7426	0.1039	1.9689
(-1, 1)	1	0.0026	0.1164	0.0499	0.5229	0.0143	0.3013	0.0563	1.3715
(-1, 0)	1	0.0054	0.2960	-0.0110	-0.1412	0.0346	0.8929	0.0431	1.2859
(0, 1)	1	-0.0076	-0.4166	0.0345	0.4427	-0.0045	-0.1179	0.0598	1.7917
(0, 2)	1	-0.0075	-0.3357	0.1059	1.1096	-0.0009	-0.0192	0.0812	1.9865
(0, 3)	1	-0.0008	-0.0310	0.0899	0.8158	-0.0187	-0.3463	0.0839	1.7775
(0, 4)	1	-0.0148	-0.5131	0.1349	1.0949	-0.0230	-0.3810	0.0753	1.4269
(0, 5)	1	-0.0171	-0.5412	0.1523	1.1284	0.0073	0.1104	0.0796	1.3770
(0, 10)	1	-0.0379	-0.8858	0.1596	0.8733	0.1250	1.3959	0.0827	1.0566
(0, 20)	1	0.0160	0.2707	0.3824	1.5145	0.1686	1.3626	0.0902	0.8340

## 5. Conclusion

We have seen how the announcement of monetary penalties related to GDPR infringement can result in (statistically significant) negative CARs of around 1% up to three days after the event. It was also observed that the economic impact on the market value of a publicly listed firm far outweighs the monetary value of the fine itself in almost all cases, and that a very small fine can have huge impact on market value (cf. BNP Paribas). We also know from the literature that CARs of a similar magnitude are generated at the time of the initial announcement of a breach. Considering all of these negative factors, the need for firms to invest in cyber security to protect data privacy is clearly underpinned by this research, as well as showing a clear economic impact of the introduction of the GDPR itself.

In light of the recent introduction of the GDPR, the dataset for this study was (necessarily) limited. Once more data becomes available and the market recovers from the COVID-19 pandemic, future research is expected to give a better idea of the impact of GDPR infringement fines on publicly listed firm value. Although four examples of GDPR fine appeals were identified and positive returns were observed where those appeals were successful (and the reverse), the results were not statistically significant, and we were unable to reject the null hypothesis of zero abnormal returns. Future research is needed in this area also – recently there has been news of Deutsche Wohnen appealing their €14.5m fine and, with the high-profile reductions of the fines for International Airlines (BA) and Marriott, a precedent appears to have been set with the ICO recognising and encouraging infringing firms to invest in cyber security measures (Macfarlanes 2020). Future studies may, therefore, reveal more about the positive impact of the GDPR on cyber security investment following its introduction and subsequent punitive actions. In this study only 2 out of 21 (10%) of ultimate parent firms were US based with the balance being European, therefore this work also begins to offset the strong US bias of these type of studies in the literature.

## Acknowledgements

The authors wish to thank the anonymous reviewers for their valuable and constructive feedback.

## References

- Andoh-Baidoo F.K., Amoako-Gyampah K., Osei-Bryson K.M. (2010), *How Internet security breaches harm market value*, *IEEE Security and Privacy* 8(1), 36–42
- BBC (2013), *Sony fined over 'preventable' PlayStation data hack*, <https://www.bbc.co.uk/news/technology-21160818>. Accessed on: 30/03/2021
- BBC (2016), *TalkTalk fined £400,000 for theft of customer details*, <https://www.bbc.co.uk/news/business-37565367>. Accessed on: 26/04/2021



- Campbell, K., Gordon, L.A., Loeb, M.P, Zhou, L. (2003), *The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market*, Journal of Computer Security, **11**(3), 431-448
- Castillo, D., Falzon, J. (2018), *An analysis of the impact of Wannacry cyberattack on cybersecurity stock returns*, Review of Economics and Finance, **13**(3), 93-100
- Cavusoglu, H., Mishra, B., Raghunathan, S. (2004), *The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers*, International Journal of Electronic Commerce, **9**, 69-104
- CMS Legal (2021), *GDPR Enforcement Tracker*, <https://www.enforcementtracker.com/>. Accessed on: 26/02/2021
- Daly A. (2018), *The introduction of data breach notification legislation in Australia: A comparative view*, Computer Law & Security Review, **34**, 477-495
- Data Protection Act (1998), <https://www.legislation.gov.uk/ukpga/1998/29/contents>. Accessed on: 30/04/2021
- Data Protection Act (2018), <http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted>. Accessed on: 10/03/2019
- Data Protection Directive (1995), <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML>. Accessed on: 30/04/21
- Deane J.K., Goldberg D.M., Rakes T.R., Rees L.P. (2019), *The effect of information security certification announcements on the market value of the firm*. Information Technology & Management, **20**(3), 107-121
- Dyckman, T., Philbrick, D., Stephan, J. (1984), *A Comparison of Event Study Methodologies Using Daily Stock Returns: A Simulation Approach*, Journal of Accounting Research, **22**, (Supplement)
- ENISA (2020), *ETL2020 The Year in Review*, <https://www.enisa.europa.eu/publications/year-in-review>
- European Commission (2021), *Proposal for an ePrivacy Regulation*, <https://digital-strategy.ec.europa.eu/en/policies/eprivacy-regulation>
- European Convention on Human Rights (1950), <https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005>. Accessed on: 30/04/21
- Fama, E. F. (1970), *Efficient Capital Markets: A Review of Theory and Empirical Work*, The Journal of Finance, **25**(2), 383-417
- Goel, S., Shawky, H.A. (2009), *Estimating the market impact of security breach announcements on firm values*, Information & Management, **46**(7), 404-410
- Goel S., Shawky, H.A., (2014) *The Impact of Federal and State Notification Laws on Security Breach Announcements*, Communications of the Association for Information Systems, **34**, 37-50
- Hinz, O., Nofer, M., Schiereck, D., Trillig, J. (2015) *The influence of data theft on the share prices and systematic risk of consumer electronics companies*, Information & Management, **52**(3), 337-347
- Jeong, C., Lee, S., Lim, J. (2019), *Information security breaches and IT security investments: Impacts on competitors*, Information & Management, **56**(5), 681-695
- Kannan, K., Rees, J., Sridhar, S., (2007), *Market Reactions to Information Security Breach Announcements: An Empirical Analysis*, International Journal of Electronic Commerce, 01 September **12**(1), 69-91
- Lin, Z., Sapp, T.R., Ulmer, J.R., Parsa, R. (2020) *Insider trading ahead of cyber breach announcements*, Journal of Financial Markets, **50**, 100527
- Macfarlanes (2020), <https://www.macfarlanes.com/what-we-think/in-depth/2020/lessons-from-the-ico-s-decisions-to-reduce-the-ba-and-marriott-gdpr-fines/>. Accessed on: 26/02/21
- Mackinlay, A. C. (1997), *Event Studies in Economics and Finance*, Journal of Economic Literature **35**(1) (March)
- Murciano-Goroff (2019), *Do Data Breach Disclosure Laws Increase Firms' Investment in Securing Their Digital Infrastructure?*, WEIS 2019
- Nieuwesteeg, B., Faure, M. (2018), *An analysis of the effectiveness of the EU data breach notification obligation*, Computer Law & Security Review: **34**(6), 1232-1246
- Privacy and Electronic Communications Directive (2002), <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32002L0058>. Accessed on: 30/04/21
- R Core Team (2018), *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available on: <https://www.R-project.org/>
- Richardson, V.J., Smith, R.E, Watson, M.W. (2019) *Much Ado about Nothing: The (Lack of) Economic Impact of Data Privacy Breaches*, Journal of Information Systems: **33**(3), 227-265
- Romanosky, S., Telang, R., Acquisti, A. (2011), *Do Data Breach Disclosure Laws Reduce Identity Theft?*, Journal of Policy Analysis and Management, **30**(2), 256-286
- Rosati, P., Deeney, P., Cummins, M., Van der Werff, L., Lynn, T. (2019), *Social media and stock price reaction to data breach announcements: Evidence from US listed companies*, Research in International Business and Finance, **47**, 458-469
- Schatz, D., Bashroush, R. (2016), *The impact of repeated data breach events on organisations' market value*, Information & Computer Security, **24**(1), 73-92
- Schimmer, M., Levchenko, A., and Müller, S. (2014), *EventStudyTools (Research Apps)*, St. Gallen. Available on: <http://www.eventstudytools.com>. Accessed on: 26/02/2021
- Spanos, G., Angelis, L. (2016), *The impact of information security events to the stock market: A systematic literature review*, Computers and Security, **58**, 216-229
- Tweneboah-Kodua, S., Atsu, F. and Buchanan, W. (2018), *Impact of cyberattacks on stock performance: a comparative study*, Information and Computer Security, **26**(5), 637-652
- Yahoo!Finance (2019), *Historical Data*, <https://finance.yahoo.com/quote>

# Side Channel Attacks and Mitigations 2015-2020: A Taxonomy of Published Work

Andrew Johnson

Faculty of Computing, Engineering and Mathematics, University of South Wales, UK

[andrew.johnson@southwales.ac.uk](mailto:andrew.johnson@southwales.ac.uk)

DOI: 10.34190/EWS.21.035

**Abstract:** Side Channel Attacks (SCAs) have become a prominent area of both research and organisational cyber defence strategies in recent years. With the advent of microprocessor performance optimizations such as speculative execution and branch prediction enhancements to Intel processors in 1996, it is possible for an adversary to target the vulnerabilities that are inherent in the optimization design. This paper presents a taxonomy of published works on the theme of hardware based SCAs and some of their mitigations from the inclusive years 2015-2020. It includes research of peer reviewed published work including open access publications. The results of research undertaken represents a large proportion of papers during the time period from select searches across three online database sources: IEEE(Institute of Electrical and Electronic Engineers); ACM (Association for Computing Machinery) Digital Library; Scopus (Elsevier/Science Direct). The taxonomy includes 684 papers from both conference and journal article publications which include SCAs, mitigations and surveys. The choice of online databases used was based on the functionality of the search engines to enable a download of search results to a 'BibTex' format for data analysis. The aim of this work is to identify and present SCAs and mitigations with two objectives: To present the published work data analysis results from the searches that show the most common and varied scope of SCA hardware targets, methods, techniques, and mitigations. To identify trends in the SCA research field over the selected time period and also present some of the more recent SCA papers that expand future research prospects.

**Keywords:** side-channel attacks, side-channel mitigations, taxonomy, hardware

---

## 1. Introduction

Recent proven exploit attacks such as Spectre (Kocher et al. 2018), Meltdown (Lipp et al. 2018) and Foreshadow (Van Bulck et al. 2018) targeting microprocessors have identified in some cases critical vulnerabilities. Also, because of the performance demands on modern Central Processing Units (CPU) due to their dramatic increase in transistor numbers yet decreased size, modern chipsets are now integrated with additional units such as Graphical Processor Units (GPU), Field Programmable Gate Arrays (FPGA) and other System on Chip (SoC) devices that ease the power consumption and resource contention from the CPU. Yet the integration of these units is not yet fully secured and existing units rely heavily on access control mechanisms as defences which are being proven to be inadequate by innovative new exploits by researchers.

In addition to CPUs, the optimization and cost reduction in manufacturing of Dynamic Random Access Memory (DRAM) has also inherited vulnerabilities. DRAM physical implementation causes individual cell proximity to be vulnerable to 'disturbance errors' on adjacent cell rows that can cause bits to be flipped on repeated access. These attacks are known as 'Rowhammer' attacks.

With improved defence mechanisms such as Intrusion Detection Systems (IDS), traditional software targeted attack methods such as malware, trojans and worms have become easier to defend against. However, the Advanced Persistent Threat (APT) of cyber attacks are becoming more complex and highly skilled attackers are finding ways of bypassing many defences by means such as SCAs on computer hardware as opposed to software based attacks only. Hence there is a need to continue research contribution into the defence of hardware focused attacks such as SCAs.

The scope of SCAs targeting computer hardware is wide and extremely varied. There are several thousand published works from its first concept in 1996 to present day, with each new published work demonstrating a new exploit that expands on previous work or circumvents mitigations of previous exploits. This ever increasing genre of cyber attacks is becoming more prevalent, and the continued exposure of inherent vulnerabilities of vendor products such as Intel's latest 10<sup>th</sup> generation microprocessors are feeding the research of cyber security in this field.

## 2. Contribution to research

The taxonomy will benefit the research and cyber security defence of SCAs by presenting some of the data retrieved from the search results of published work over the last six years 2015-2020 (inclusive). It will provide a means of reference of demonstrated attack vectors, methods and techniques used in SCAs on computer hardware will assist in the identification of areas of research that can be further explored to identify new attack vectors, vulnerabilities, and exploits.

## 3. Similar works

Tsalis et al. (2019) published a paper which provides a taxonomy of SCAs on Critical Infrastructures (CI). Their categorization is valuable in that it categorizes attacks based on the SCA category, hardware or software target and the resulting data exfiltrated. Sayakkara et al. (2019) also recently completed a survey of EM SCAs from a digital forensic perspective. A few other valuable SCA surveys over the last five years include Lyu and Mishra (2017) presenting 'A Survey of Side-Channel Attacks on Caches and Countermeasures'; Szefer (2019) presenting 'Survey of Microarchitectural Side and Covert Channels , Attacks , and Defenses' and work by Spreitzer et al. (2018) on the 'Systematic Classification of Side-Channel Attacks: A Case Study for Mobile Devices'.

## 4. Literary searches

The most common SCA types were searched for in Scopus (Elsevier/Science Direct), IEEE Xplore and ACM Digital databases during the time period of 2015-2020. A total of 684 papers are represented in this taxonomy across the time period. The search string format used is shown below.

### 4.1 Search string

Search question: Find papers containing the words 'side channel attack' and 'hardware' and also 'power analysis' or 'cache timing' or 'electromagnetic (EM) analysis' or 'microarchitectural' or 'speculative execution'.

Search String: ( side AND channel AND attacks ) AND hardware AND ( power AND analysis OR cache AND timing OR electromagnetic AND analysis OR microarchitectural OR speculative AND execution )

### 4.2 Research question

The research should answer the following question: What are the most common SCA hardware targets, types, methods, techniques, and mitigations used over the last six years?

## 5. Data analysis – SCA publications by year 2015-2020

The following graph shows the results of the data search by numbers of SCA publications from 2015-2020. There is a steady increase in the numbers over the time period. Factors that have contributed to the rise in 2018 suggest not only an increase in found vulnerabilities and exploit variants through research, but also the financial incentive behind finding new vulnerabilities by reward systems such as Intel's 'Bug Bounty Program'(Intel Corporation 2019) introduced in 2017 after the publication of the Spectre exploit.

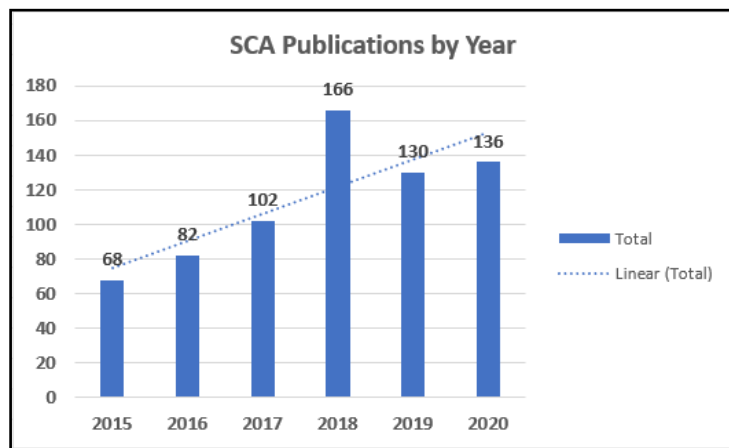


Figure 1: SCA publications by year

### 5.1 Hardware SCA targets 2015-2020

From the data retrieved from the searches, Figure 2 identifies the predominant targeted hardware demonstrated in the peer reviewed papers. As can be seen, researchers have targeted many hardware vectors across various devices. The top five hardware targets are discussed in the next sections literature review, with brief reviews of some of the smaller representation of hardware targets.

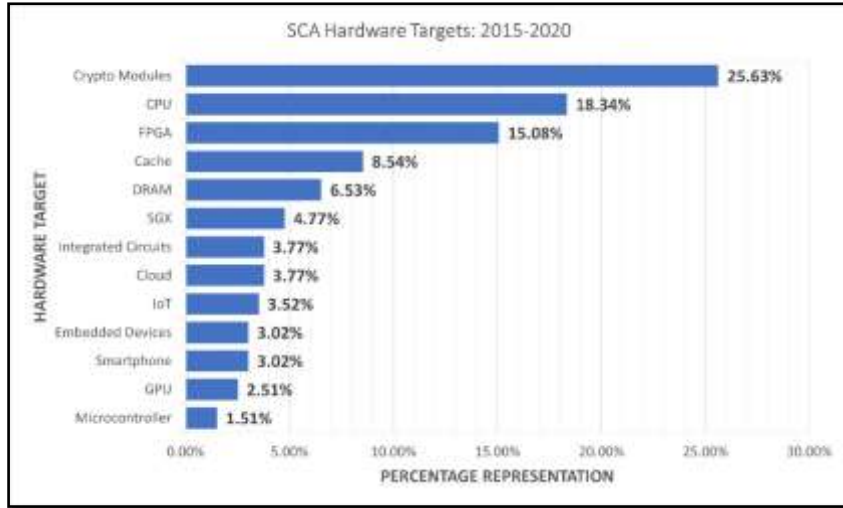


Figure 2: SCA hardware targets 2015-2020

### 5.2 Hardware target descriptions

To provide an all-encompassing overview and critical analysis of every represented hardware attack vector on computer systems is beyond the scope of this paper. A computer system is not only complex but has potentially hundreds of vulnerabilities inherent in its hardware across multiple different vendors and platforms. The most frequently used hardware targets from the literary search is described below.

#### 5.2.1 Cryptographic modules (~25%)

The data from the searches show that cryptographic modules continue to be the main research SCA vector. The term ‘cryptographic module’ is not specifically related to a hardware component. The modules can be solely hardware based, software based or a combination of both. In addition, they can be part integrated as part of the CPU design, or lie elsewhere in hardware such as FPGA’s. Researchers have identified an abundance of means to extract private keys during the encryption process via many SCA methods. The scope of variation in algorithms used in the encryption process is also demonstrated in the published works. The SCA methods used to retrieve private keys is also diverse and ranges from frequently used techniques such as power analysis, timing analysis, EM analysis to more uncommon methods such as deep learning analysis, reverse engineering and instruction set translations. The general term used to describe these attacks is ‘cryptographic analysis’.

#### 5.2.2 CPU – functional/logical side channels (~19%)

In the context of microprocessors or CPU’s, the introduction of performance enhancing functionality led to an identification of inherent vulnerabilities. Intel describe the P6 microarchitecture of its CPUs as a: ‘three-way superscalar, pipelined architecture’, which can ‘decode, dispatch and complete execution of three instructions per clock cycle’, (Intel 2006) which dramatically increases the performance of the processing speed by execution ‘out of order’. Dynamic Execution involves three concepts: ‘Deep Branch Prediction, Dynamic Data Flow Analysis, Speculative Execution’ (Intel 2006). It is the speculative execution and branch prediction functionality that is exploited by Spectre and Meltdown attacks.

#### 5.2.3 Field Programmable Gate Array (FPGA) (~15%)

FPGAs have the flexibility to be programmable for specific tasks required by developers. One of the vulnerabilities is the FPGA on chip power monitors that can be used as a side-channel. Where FPGAs are multi tenants, the internal on-chip power monitor units emit a ‘cross talk frequency’ between the wires of their co-

tenants that can leak information. Ramesh et al. (2018) show experiments in their paper *'FPGA Side Channel Attacks without Physical Access'* require no physical access to the device and attack the Intel Static Random Access Memory (SRAM) FPGA using the chip monitor software to conduct Differential Power Analysis (DPA) and frequency measurements across the long wires of the FPGA's. They successfully extract a full AES encryption key using this attack. Zhao and Suh (2018) conduct an almost identical attack in their paper. They similarly attack the on-chip monitors to conduct remote power analysis SCAs using Simple Power Analysis (SPA) techniques. They also demonstrate how the FPGA can be exploited to conduct cross FPGA to CPU exploits.

#### 5.2.4 Caches (~9%)

CPUs have a cache hierarchy built into the same chipset that houses the processor. These caches are typically layered onto the chipsets with three to four caches. L1 cache being the closest proximity to the CPU and hence the fastest, then L2/3 and Last Level Caches (LLC) being furthest away from the CPU. The closer the cache, the faster the data retrieval process during execution. This optimization has clear advantages when the data in the cache is used or taken instead of the CPU having to wait for data from main memory. However, vulnerabilities in caches have been widely exposed by researchers. Particularly, when instructions have been speculatively executed and the resulting data is cached but not used. Cache exploitation is probably the most extensively researched and reported vulnerability on the theme of SCAs. Although Figure 2 only shows a ~9% representation, it should be noted that some attacks on other hardware vectors such as CPU and cryptographic modules include methods that utilise the cache. For example, a majority of the SCAs use some means of cache timing analysis to retrieve secret data. Kocher (1996) demonstrated cache timings on RAM caches to conduct cryptographic analysis. In addition, the papers have subsequently followed seminal works by Osvik et al. (2006) whose *'Cache Timing Attacks and Countermeasures: The case of AES'* describes the 'Evict & Time' and 'Prime & Probe' cache timing attack methods to also extract cryptographic keys; 'Flush & Reload' by Yarom and Falkner (2014) and 'Evict & Reload' by Gruss et al. (2015) with many latter variants demonstrated in the researched papers since 2015.

#### 5.2.5 Dynamic Access Memory (DRAM) (~7%)

The 'Rowhammer' exploit presented by Kim et al. (2014) in their paper *'Flipping bits in memory without accessing them: An experimental study of DRAM disturbance errors'* led to an expansion in published works post 2014 on the theme of Rowhammer SCAs. In 2018 alone, there were well over thirty publications on Rowhammer attacks. Rowhammer has become the academic publication of choice in regard to EM disturbance side-channel vulnerabilities in DRAM. The Rowhammer exploit reads memory addresses contained in DRAM rows repeatedly. The flushing of the cache is required to ensure the DRAM rows are read during each iteration as opposed to just the caches being read. Hence with repeated iterations or 'hammering' of the DRAM memory rows, eventually adjacent row cells can be flipped caused by the cell 'disturbance'. Examples of novel work in the data include *'Dedup Est Machina: Memory Deduplication as an Advanced Exploitation Vector'* (Bosman et al. 2016), *'Nethammer: Inducing Rowhammer Faults through Network Requests'* (Lipp et al. 2018b) and *'Exploiting Correcting Codes: On the Effectiveness of ECC Memory Against Rowhammer Attacks'* (Cojocar et al. 2019),

#### 5.2.6 IoT/integrated circuits/embedded devices (combined ~10%)

Although represented separately in Figure 2 due to their different terminology, these targets are essentially the same. Encapsulating the terminology into simply 'IoT' devices is preferable and within this group there are potentially hundreds if not thousands of devices from smart home fridges to home security cameras. Hence there has been a lot of research conducted across the field. An example of a survey include *'SoK: Security Evaluation of Home-Based IoT Deployments'* (Alrawi et al. 2019), that provides a comprehensive security evaluation of IoT devices and vendor targets. The paper surveys many vulnerabilities of IoT devices and highlights the lack of integrated security on core devices that can cause side-channel leakage that can be exploited.

#### 5.2.7 Cloud/machine virtualization(shared hardware)(~4%)

Although VM's are software based, it is worth mentioning the attacks on VM's from cross tenants sharing the same hardware, especially in the context of attacks such as Foreshadow-NG that exploit the boundaries of VM shared hardware resources. Papers have used attack techniques such as Rowhammer across shared host VM,s. Cross VM attacks have been demonstrated by *'A survey on the security of hypervisors in cloud computing'* (Riddle and Chung 2015), *'Flip Feng Shui: Hammering a Needle in the Software Stack'* (Razavi et al. 2016), *'One Bit Flips,*

*One Cloud Flops: Cross-VM Row Hammer Attacks and Privilege Escalation* (Xiao et al. 2016) and *Foreshadow-NG: Breaking the Virtual Memory Abstraction with Transient Out-of-Order Execution* (Weisse et al. 2018).

### 5.3 Other hardware represented

Other hardware targets presented in the data with examples of authors include:-

- Memory Management Units (MMU) - ‘Malicious Management Unit: Why Stopping Cache Attacks in Software is Harder Than You Think’. (Van Schaik et al. 2018)
- Graphical Processor Units (GPUs) – ‘Grand Pwning Unit: Accelerating Microarchitectural Attacks with the GPU’ (Frigo et al. 2018)
- Power Management Units (PPUs) - PMU-Trojan: ‘On exploiting power management side channel for information leakage’ (Islam and Kundu 2018)
- Smartphones - ‘Nethammer: Inducing Rowhammer Faults through network requests’ (Lipp et al. 2018)
- Hard Disk Drives (HDDs) – ‘Hard drive side-channel attacks using smartphone magnetic field sensors.’ (Biedermann et al. 2015)
- Near Field Communication Devices (NFCs) – ‘Organisational Aspects and Anatomy of an Attack on NFC/HCE Mobile Payment Systems’(Cavallari et al. 2015)
- Smartwatches – ‘Side-Channel Inference Attacks on Mobile Keypads Using Smartwatches.’ (Maiti et al. 2018)
- NAND Flash Memory – ‘Read Disturb Errors in MLC NAND Flash Memory: Characterization, Mitigation, and Recovery’ (Cai et al. 2015).

## 6. SCA types 2015-2020

SCA types are the source of where the computer by-product or leakage comes from. An example being EM energy which is a by-product of current running through an electronic circuit that can be measured via its frequency traces. Figure 3 below demonstrates the most common SCA types presented from the data taken from the literary searches. From Figure 3 it can be seen that the power usage of a computer system has been the side channel type of choice for researchers over the last six years. In fact power analysis was first demonstrated in *‘Differential Power Analysis’* by Kocher et al. (1999), which presented the first power based SCA on cryptographic implementations. It should be noted that the graph below does not demonstrate combined side channel attack types, such as *‘Physical key extraction attacks on PCs’* by Genkin et al. (2016) that demonstrates SCAs via power analysis, EM analysis and acoustic inference, but more so a representation of the most frequently used side channel types.

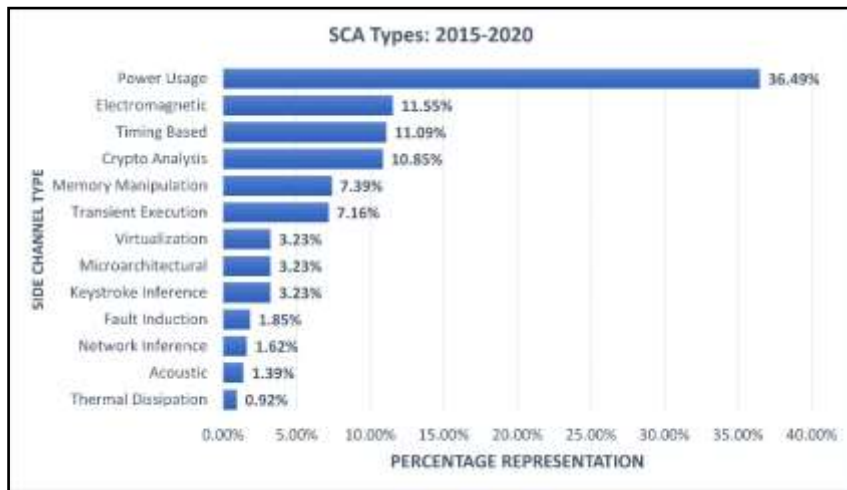


Figure 3: SCA types: 2015-2020

## 7. SCA methods 2015-2020

A side-channel attack method is categorized by the method used to analyse the SCA Types as seen in Figure 4 below. For example, power analysis is used to detect and analyse power traces given off by a particular device or operation of a computer system during its power usage. From Figure 4 it can be seen that power analysis is

the most heavily researched, making up almost 37% of published work, with cache timing, EM and cryptographic analysis taking up a combined ~34%. As in the graph shown in Figure 2, the graph below also does not include a breakdown of combined attack methods but more a breakdown of all side-channel attack methods by percentage of the frequency of their appearance across the published works.

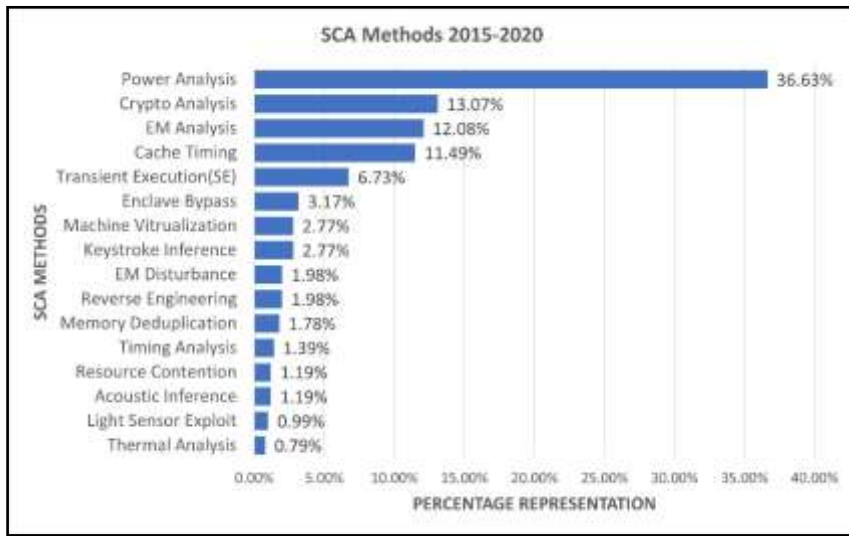


Figure 4: SCA methods 2015-2020

### 8. SCA techniques 2015-2020

A side channel technique is categorized as a ‘sub-method’ of a SCA method. The diversity of SCA techniques is shown in well over a hundred different techniques represented in the data. Figure 5 shows that power analysis techniques such as DPA, SPA and CPA (Correlation Power Analysis) have a combined contribution of ~30% of all SCA techniques used. One of the most diverse attack techniques stems from the Cache Timing SCA method which has an increasing number of techniques researched such as ‘Flush & Reload’, a cache timing analysis technique used in both Spectre and Meltdown exploits and many more subsequent works.

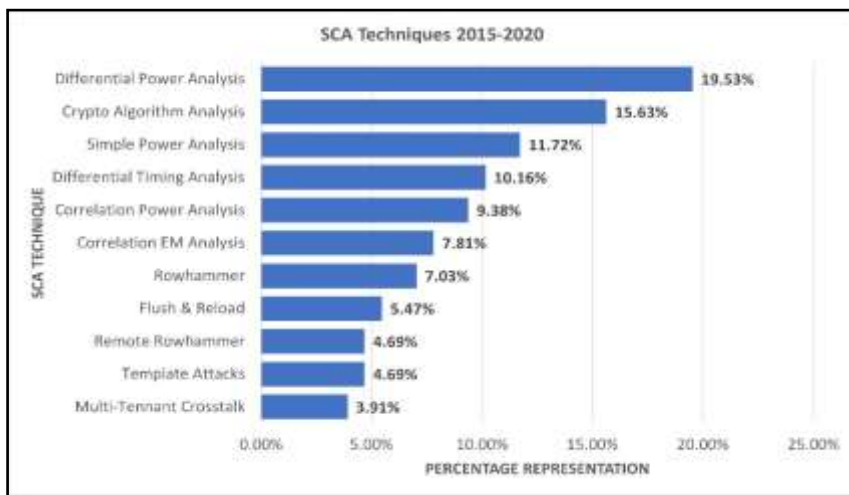


Figure 5: SCA techniques 2015-2020

### 9. SCA mitigations 2015-2020

The diversity seen in published works in regard to SCA types, methods and techniques can also be seen in the variety of mitigation techniques used to remove or reduce the risk of exploiting SCA vulnerabilities as well as using side channels as protection mechanisms to monitor computer systems. Figure 6 below shows the most commonly used mitigation techniques.



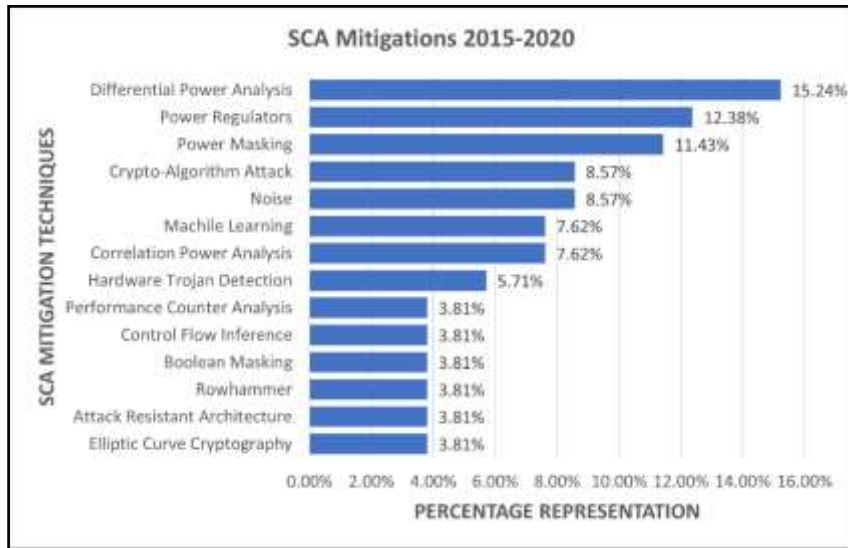


Figure 6: SCA mitigations 2015-2020

### 9.1 Masking (~14%)

Masking is predominantly used to mitigate power analysis techniques such as DPA. Masking is quite a generic term but essentially means hiding the characteristics of particular side channel such as power usage so that adversaries cannot use power analysis SCA methods and techniques. One technique of masking is to ‘mask the parameters used in the cryptographic process’ as introduced by Kocher et al. (2011). This technique is also known as ‘blinding’.

### 9.2 Power regulators (~12%)

On-chip power regulators can be used to control measurable power differences between different instructions that make power analysis techniques more difficult, particularly DPA that relies on significant power differentiation in order to infer the instructions and data being processed. Mitigations have been demonstrated by Yu et al. (2015) and Singh et al. (2015).

### 9.3 Error Correcting Code (ECC) Implementation (~4%)

ECC involves implementing extra ‘bit checks’ onto DRAM chip modules that prevent bit flipping caused by exploits such as Rowhammer. However there is evidence that ECC is limited to defend multiple cell Rowhammer attacks. Cojocar et al. (2019) have recently published a study on the inadequacies of ECC in their paper entitled ‘Exploiting Correcting Codes: On the Effectiveness of ECC Memory Against Rowhammer Attacks’ that demonstrates how conducting a ‘cold-boot’ attack and reverse engineering the ECC functions can be used to further attack the ECC memory controller to conduct advanced Rowhammer. They call their exploit ‘ECCploit’ which renders ECC implementations defenceless.

### 9.4 Noise injection (~9%)

Noise injection involves the addition of additional signal transmissions into a computer system that can make side-channel analysis such as power and EM analysis more difficult as many of the attack types involve the analysis of power or EM signals through trace captures that can be analysed through devices such as oscilloscopes. With the addition of noise into the signals the traces become more complex or even impossible to analyse. A good example of work is presented in ‘ASNI: Attenuated Signature Noise Injection for Low-Overhead Power Side-Channel Attack Immunity’ (Das et al. 2018)

### 9.5 Machine learning (~8%)

Several papers have presented Machine Learning (ML) techniques to defend against crypto analysis based attack methods. Shan et al. (2017) present a ‘machine learning trained power consumption module’ that disrupts the key extraction process of the Hamming Distance (HD) cryptographic algorithm attacks. Mushtaq et al. (2018)



have also used ML to protect against cache based ‘Flush & Reload’ timing based attack techniques. Machine learning and ‘deep learning’ are the novel science being used for SCA mitigation research.

### 10. Seminal works pre-2015

Table 1 below represents some of the most cited published works on the theme of SCAs, including the seminal works of Paul C Kocher. Although this list is not inclusive and by no means a representative ‘top ten’ of works on the subject, the papers are a good representation of their particular field. Seminal works include Osvik et al. (2006) who were one of the first to show AES key extraction through cache timing analysis and Yarom and Falkner (2014) who introduced the concept of the ‘Flush & Reload’ cache timing analysis technique.

**Table 1:** Most cited published works

Year	Title	SC Type	Citations	Ref
1996	Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems	Timing	4642	Kocher, P.C. (1996)
1999	Differential Power Analysis	Power Usage	7373	Kocher et al. (1999)
2005	Side-channel attacks: Ten years after its publication and the impacts on cryptographic module security testing	Survey	230	Zhou and Feng (2005)
2006	Cache attacks and counter-measures: The case of AES	Timing	920	Osvik et al. (2006)
2006	Covert and side channels due to processor architecture	Microarchitectural	216	Wang and Lee (2006)
2011	Introduction to differential power analysis	Power Usage	374	Kocher et al. (2011)
2014	Flush + Reload : a High Resolution, Low Noise, L3 Cache Side-Channel Attack	Timing	550	Yarom and Falkner(2014)
2014	Flipping bits in memory without accessing them: An experimental study of DRAM disturbance errors	EM	398	Kim et al. (2014)
2014	TouchLogger: Inferring Keystrokes on Touch Screen from Smartphone Motion	Keystroke inference	544	Nilsson et al. (2014)
2014	Cross-Tenant Side-Channel Attacks in PaaS Clouds	EM	214	Zhang et al. (2014)

### 11. Future trends - Microarchitectural Data Sampling (MDS)

In May 2019, ‘Transient Execution’, a derivative of speculative execution was demonstrated in attacks by the ‘VUsec’ department of the Vrije University in Amsterdam. The first of these called ‘Fallout’ by Minkin et al. (2019) has circumvented the mitigations for meltdown implemented in the processor hardware of Intel’s 9<sup>th</sup> generation processors by exploiting Write Transient Forwarding (WTF), the process by which CPU virtual to physical address mappings are made via a ‘store buffer’. Van Schaik et al. (2019) also released a paper in May 2019 which further exploits the address space resolution and privilege boundary mechanisms. They call their exploit ‘*Rogue In-flight Data Load (RIDL)*’. The two papers described above are examples of a ‘new wave’ of hardware SCAs research emerging known as ‘*Microarchitectural Data Sampling*’ (MDS) by the Vrije Universiteit Amsterdam (2019). The most recent to emerge from the VUsec department is ‘*CrossTalk: Speculative Data Leaks Across Cores Are Real*’ (Ragab et al. 2021) which has been accepted for publication in 2021. CrossTalk demonstrates even more advanced research into extraction of data from within ‘staging buffers’ that transfer data between CPU cores.

### 12. Conclusion

This paper has provided a taxonomy of some of the peer reviewed published papers from 2015-2019 on the theme of SCAs and their mitigations. The study included the search and resulting data analysis of 684 papers from the time period from three online database resources. The data was placed into five main categories: SCA Types; SCA Hardware Targets; SCA Methods; SCA Techniques; SCA Mitigations. The resulting data was presented in pivot chart form with descriptions of some of the SCA categories and consequent brief literature reviews. By providing this data, the author has provided some further clarity in the field of SCA research from an academic perspective through a taxonomy.

### References

Alrawi, O. et al. (2019) SoK: Security Evaluation of Home-Based IoT Deployments. In: *Proceedings - IEEE Symposium on Security and Privacy*.

Biedermann, S., Katzenbeisser, S. and Szefer, J. (2015) Hard drive side-channel attacks using smartphone magnetic field sensors. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

Bosman, E. et al. (2016) Dedup Est Machina: Memory Deduplication as an Advanced Exploitation Vector. In: *Proceedings - 2016 IEEE Symposium on Security and Privacy, SP 2016*.

Van Bulck, J. et al. (2018) Foreshadow: Extracting the Keys to the Intel SGX Kingdom with Transient Out-of-Order Execution. *Proceedings of the 27th USENIX Security Symposium*

Cai, Y. et al. (2015) Read Disturb Errors in MLC NAND Flash Memory: Characterization, Mitigation, and Recovery. In: *Proceedings of the International Conference on Dependable Systems and Networks*.

Cavallari, M., Adami, L., Tornieri, F. (2015) Organisational Aspects and Anatomy of an Attack on NFC/HCE Mobile Payment Systems.

- Cojocar, L. et al. (2019) Exploiting Correcting Codes: On the Effectiveness of ECC Memory Against Rowhammer Attacks. In: *SP - IEEE Symposium on Security and Privacy*.
- Das, D. et al. (2018) ASNI: Attenuated Signature Noise Injection for Low-Overhead Power Side-Channel Attack Immunity. *IEEE Transactions on Circuits and Systems I: Regular Papers*.
- Frijo, P. et al. (2018) Grand Pwning Unit: Accelerating Microarchitectural Attacks with the GPU. In: *Proceedings - IEEE Symposium on Security and Privacy*.
- Genkin, D. et al. (2016) Physical key extraction attacks on PCs. *Communications of the ACM* 59(6), pp. 70–79.
- Gruss, D., Spreitzer, R. and Mangard, S. (2015) Cache template attacks: Automating attacks on inclusive last-level caches. *Proceedings of the 24th USENIX Security Symposium*
- Intel (2006) Intel® 64 and IA-32 Architectures Software Developer Manuals.[online] <https://software.intel.com/en-us/articles/intel-sdm>
- Intel Corporation (2019) Bug Bounty Program.[online] <https://www.intel.com/content/www/us/en/security-center/bug-bounty-program.html>
- Islam, M.N. and Kundu, S. (2018) PMU-Trojan: On exploiting power management side channel for information leakage. In: *Proceedings of the Asia and South Pacific Design Automation Conference, ASP-DAC*.
- Kim, Y. et al. (2014) Flipping bits in memory without accessing them: An experimental study of DRAM disturbance errors. In: *Proceedings - International Symposium on Computer Architecture*. IEEE.
- Kocher, P., Jaffe, J. and Jun, B. (1999) Differential Power Analysis. *Journal of Cryptographic Engineering*, pp. 388–397.
- Kocher, P. et al. (2011) Introduction to differential power analysis. *Journal of Cryptographic Engineering* 1(1), pp. 5–27.
- Kocher, P. et al. (2018) Spectre Attacks: Exploiting Speculative Execution. *40th IEEE Symposium on Security and Privacy (S&P'19)*.
- Kocher, P.C. (1996) Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. *Crypto*, pp. 104–113.
- Lipp, M. et al. (2018a) Meltdown: Reading Kernel Memory from User Space. *27th USENIX Security Symposium (USENIX Security 18)*.
- Lipp, M. et al. (2018b) Nethammer: Inducing Rowhammer Faults through Network Requests. *ArXiv*.
- Lyu, Y. and Mishra, P. (2017) A Survey of Side-Channel Attacks on Caches and Countermeasures. *Journal of Hardware and Systems Security*, pp. 33–50.
- Maiti, A. et al. (2018) Side-Channel Inference Attacks on Mobile Keypads Using Smartwatches. *IEEE Transactions on Mobile Computing*.
- Minkin, M. et al. (2019) Fallout: Reading kernel writes from user space. *arXiv*
- Mushtaq, M. et al. (2018) Machine Learning For Security: The Case of Side-Channel Attack Detection at Run-time. In: *Proceedings of the 25th IEEE International Conference on Electronics, Circuits, and Systems, ICECS-2018*.
- Osvik, D.A., Shamir, A. and Tromer, E. (2006) Cache attacks and counter-measures: The case of AES. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Ragab, H. et al. (2021) CrossTalk: Speculative Data Leaks Across Cores Are Real. In: *IEEE Symposium on Security & Privacy 2021*.
- Ramesh, C. et al. (2018) FPGA Side Channel Attacks without Physical Access. In: *Proceedings - 26th IEEE International Symposium on Field-Programmable Custom Computing Machines, FCCM 2018.*, pp. 45–52.
- Razavi, K. et al. (2016) Flip Feng Shui: Hammering a Needle in the Software Stack. In: *Proceedings of the 25th USENIX Conference on Security Symposium*.
- Riddle, A.R. and Chung, S.M. (2015) A survey on the security of hypervisors in cloud computing. In: *Proceedings - 2015 IEEE 35th International Conference on Distributed Computing Systems Workshops, ICDCSW 2015*.
- Sayakkara, A., Le-Khac, N.A. and Scanlon, M. (2019) A survey of electromagnetic side-channel attacks and discussion on their case-progressing potential for digital forensics. *Digital Investigation*.
- Shan, W. et al. (2017) Machine learning based side-channel-attack countermeasure with hamming-distance redistribution and its application on advanced encryption standard. *Electronics Letters*.
- Singh, A. et al. (2015) Exploring power attack protection of resource constrained encryption engines using integrated low-drop-out regulators. In: *Proceedings of the International Symposium on Low Power Electronics and Design*.
- Spreitzer, R. et al. (2018) Systematic Classification of Side-Channel Attacks: A Case Study for Mobile Devices. *IEEE Communications Surveys and Tutorials (2018) 20(1)* 20(1), pp. 465–488.
- Szefer, J. (2019) Survey of Microarchitectural Side and Covert Channels, Attacks, and Defenses. *Journal of Hardware and Systems Security (2019) 3(3)*, pp. 219–234.
- Tsalis, N. et al. (2019) A Taxonomy of Side Channel Attacks on Critical Infrastructures and Relevant Systems. In: *Advanced Sciences and Technologies for Security Applications*. Springer, pp. 283–313.
- Van Schaik, S. et al. (2018) Malicious Management Unit: Why Stopping Cache Attacks in Software is Harder Than You Think. In: *Usenix Security 2018*.
- Van Schaik, S. et al. (2019) RIDL: Rogue in-flight data load. In: *Proceedings - IEEE Symposium on Security and Privacy*. doi: 10.1109/SP.2019.00087.
- Vrije Universiteit Amsterdam (2019) RIDL and Fallout: MDS attacks.[online] <https://mdsattacks.com/>
- Weisse, O. et al. (2018) Foreshadow-NG: Breaking the Virtual Memory Abstraction with Transient Out-of-Order Execution. *White Paper*

**Andrew Johnson**

- Xiao, Y. et al. (2016) One Bit Flips, One Cloud Flops: Cross-VM Row Hammer Attacks and Privilege Escalation. In: *25th USENIX Security Symposium*
- Yarom, Y. and Falkner, K. (2014) Flush + Reload : a High Resolution , Low Noise , L3 Cache Side-Channel Attack. In: *Proceedings of the 23th USENIX Security Symposium*.
- Yu, W., Uzun, O.A. and Kose, S. (2015) Leveraging on-chIP voltage regulators as a countermeasure against side-channel attacks. In: *Proceedings - Design Automation Conference*.
- Zhao, M. and Suh, G.E. (2018) FPGA-Based Remote Power Side-Channel Attacks. In: *Proceedings - IEEE Symposium on Security and Privacy*.

# Sanctions and Cyberspace: The Case of the EU's Cyber Sanctions Regime

Eleni Kapsokoli

University of Piraeus, School of Economics, Business and International Studies,

Department of International and European Studies, Piraeus, Greece

Laboratory of Intelligence and Cyber-Security

[ekapsokoli@unipi.gr](mailto:ekapsokoli@unipi.gr)

[elenikapsokoli1989@gmail.com](mailto:elenikapsokoli1989@gmail.com)

DOI: 10.34190/EWS.21.029

**Abstract:** Over the past years, the European Union has faced a number of cybersecurity challenges that range from cyberattacks to the critical infrastructure to cases of ransomware. In order to face these security challenges, the EU has developed all the necessary strategies and policies, including the launch of the Cyber Diplomacy Toolbox in 2017. This toolbox includes among others, the policy instrument of cyber sanctions (Council Decision 2019/796 and Council Regulation 2019/797). The purpose of this paper is to review the EU's cyber sanctions regime. In order to do that, we will apply key questions that have arisen in the sanctions literature and apply them in the case of the EU. In particular, we will examine the technical, political and judicial parameters, which involve the implementations of cyber sanctions. Issues like the reliable attribution of cyberattacks, the clarifications of relevant norms of responsible state behavior in cyberspace, the level of cooperation with the private sector and the scale and type of cyber sanctions are only some of the factors that will determine the success of the cyber sanctions regime. Having established a clear theoretical framework on the implementation of cyber sanctions, we will briefly review the empirical evidence, which involves the sanctions package that the EU announced on 17 May 2019 and the amending version published on 30 July 2020, against various entities and individuals. The end goal is to reach a conclusion on whether cyber sanctions are of symbolic nature, or can be considered as an effective policy instrument for the EU. This paper highlights the uses and limits of the EU cyber sanctions regime, which is a rather recent development and therefore under developed in the relevant literature.

**Keywords:** sanctions, cybersecurity, cyber sanctions, attribution, European Union

---

## 1. Introduction

Cyberspace is a domain where norms and international regulations are gradually defining what is permissible and what is not. The increasing number of cyber incidents over the past decade, demonstrates how state and non-state actors have developed a wide range of malicious cyber capabilities that include among others cyberattacks against the critical infrastructure, the dissemination of fake news and propaganda and the use of malware. Both state and non-state actors use cyberspace in order to achieve their political goals. Malicious activities in cyberspace have become a growing threat and therefore, the European Union (EU) and its member-states have developed a number of institutions and policies in order to secure cyberspace and its users.

Among others, the EU adopted a cyber diplomacy toolbox which contains a number of diplomatic and operational measures, including the use of sanctions. In particular, the Council Regulation 2019/796 and the Council Decision 2019/797 established a cyber sanctions regime. The cyber sanctions regime is a joint action of the EU which aims to secure a free and open cyberspace, based on norms and international law. Apart from the US, the EU is the only international actor that has proposed and applied cyber sanctions as a policy option in order to deal with cybersecurity issues (Pawlak & Biersteker, 2019, 9).

Bearing in mind that cyberspace is a domain that affects every facet of politics and security, it is critical to investigate how the concept of sanctions applies in this new and complex domain. Thus, this paper aims to evaluate the EU's cyber sanction regime. In order to do that, we will first review the concept of sanctions. Having established a solid understanding of sanctions, we will analyze the EU policies and instruments that relate to the establishment of a cyber sanction regime. The empirical evidence from the cyber sanctions imposed to Chinese, Russian and North Korean actors, will enable us to reach safer conclusions about the utility of cyber sanctions.

## 2. On sanctions

The term sanction is one of the most complicated and confusing terms in international politics. The international community is using sanctions to change the behavior of a country or regime, in cases where that country or regime is violating human rights, waging war or endangering international peace and security (Baldwin, 1985).

Sanctions is a tool or a mean for the policymakers to impose specific norms and behaviour to another actor. Sanctions can be of economic nature, but also include restrictions like travel bans and arms embargo. The sanction can be defined as “an economic instrument which is employed by one or more international actors against another, ostensibly with a view to influencing that entity’s foreign and/or security policy behavior” (Taylor, 2010, 12). International actors use sanctions not only to shape or influence a state or generally an actor’s behavior or to disregard publicly or to put pressure to other state or non-state actors, but also to influence other actors in order to support their policies and strategies.

For some scholars and practitioners, sanctions are mainly symbolic, they encapsulate the desire to act, but are not always effective in terms of shaping or influencing the behaviour of the targeted actors. In many cases, sanctions are imposed as a result of external pressure. In addition, there are also difficulties in applying such restrictive measures, since some countries are unwilling to participate. Furthermore, the targeted actors can diversify their policies regarding the sectors that are targeted (e.g. diversify their production, import from other sources, etc.) and thereby become more independent and minimise the impact of sanctions (Taylor, 2010, 19). Thus, sanctions are perceived as symbolic actions - when other diplomatic tools have failed or are unavailable - that facilitate the international or domestic pressure for responsible behaviour (Taylor, 2010, 20). Finally, the imposition of sanctions can have collateral damages, not only for the targeted state or non-state actors, but also for the actors operating in the perpetrator’s environment (Pawlak & Biersteker, 2019, 7).

In general, it is difficult to value the effectiveness of sanctions. Although there are cases in the past, where sanctions have proved to be unsuccessful, this does not exclude the possibility that they may be more useful in the future. There are two points that we should consider when applying sanctions. First, sanctions’ efficiency must be seen in terms of the alternative solutions available to the policy maker at any given situation. In many cases where sanctions were imposed, they were not one of the many available options, but rather the only option. Second, sanctions have proved to be successful when they are multilateral and proportional to the goal to be achieved. Unless these conditions are met, sanctions will be counterproductive and weaken the credibility of those who impose this policy.

### **3. Cyber sanctions and attribution**

Cyber sanctions aim to deter and respond to malicious activities in cyberspace. Cyber sanctions may be public or private and must be the result of attribution by a state. Attribution is when a state or a company accuses another actor for an attack publicly. In order to attribute a cyberattack effectively, one must reveal the computer and network systems that were responsible for the attack and identify also the individuals that were responsible for these systems (Ivan, 2019, p. 8). The cyber sanctions are difficult to be imposed because the collection of the necessary intelligence to attribute a cyberattack is a challenging task.

Cyber attackers, use crypto links, zombie routers and other ways to safeguard their anonymity and their location. The use of IP addresses as evidence of a cyberattack is not sufficient, because the attacker can alter or hide the location of the IP address. The location of the perpetrator’s IP address is not solid evidence, due to its ability to protect its identity. Thus, locating the origin of the attacker is a major security issue. The cyber attacker covers its affiliation with state actors, in order to protect its source and acts as a proxy via false-flag cyberattacks. The rise and spread of information and communication technologies and emergence of smart cities facilitates the purposes of future perpetrators. Hard evidence on attribution, requires information exchanges regarding the nature of the attacks, its actors and the vulnerabilities of the critical infrastructures. Posing sanctions based on inaccurate and non-credible evidence could cause a diplomatic and political incident (Pawlak & Biersteker, 2019, 88).

A question that is inevitable raised is whether cyber sanctions will be an effective tool to deter and respond to threats and attacks in cyberspace. So far the record is rather poor, since the policy option of cyber sanctions has rarely been used. The US government was the first country who imposed cyber sanctions against North Korea in 2015 as a response of an alleged cyberattack on Sony Pictures Entertainment (Liaropoulos, 2018, 265). So far, the US government has imposed cyber sanctions against North Korean officials, Russian intelligence services (the Main Intelligence Directorate- GRU and the Federal Security Service- FSB) and three companies that supported the cyber activities of GRU (Moret & Pawlak, 2017, 2). Bearing in mind the relative effectiveness of sanctions in general, the difficulties in attributing cyberattacks and the broader policy options that are available, we will now review the cyber sanctions regime established by the EU.

#### **4. The EU's sanction regime for cyberattacks**

The EU is no stranger to sanctions as a policy instrument. Sanctions involve the preventive measures (capacity building, awareness raising), the cooperative measures (dialogues, demarches), the stabilizing measures (statements, council conclusions, demarches, dialogues), the restrictive measures (asset freezes, travel bans) and the supportive measures (lawful responses with the use of article 51 or 42 paragraph 7 of the EU) (European Union External Action, 2016). As part of the Common Foreign and Security Policy, sanctions include interruption of diplomatic relations, recall of diplomatic representatives, arms embargoes, restrictions on admission (travel bans), freezing of assets and economic sanctions or restrictions (Pawlak & Biersteker, 2019, 8). Over the past years, the EU has imposed 37 sanctions regimes. Some of these cases include the countering of terrorism in Libya in 1999, the case of Al Qaeda, the support of democracy, human rights and the law in the case of Belarus in 2006, and in relation to conflict management in the cases of Libya and Syria since 2011 (Moret & Pawlak, 2017, 2). The use of sanctions in the above cases was the desirable tool by policymakers when diplomacy seemed as a less effective strategy.

In terms of cybersecurity, the EU has developed over the last years a number of policies, strategies and institutions, in order to safeguard its member-states in cyberspace. The EU aims to strengthen resilience in cyberspace, to build trust, to prevent conflicts, to protect human rights and freedoms and to promote multilateralism. In order to achieve the above, the EU needs to develop the necessary capabilities and technical know-how, to promote relevant norms and to engage both the civil society and the private sector in governing cyberspace (Kapsokoli, 2020).

It is in this context that the European Council, adopted on 19 June 2017 a framework for a joint diplomatic response to malicious cyber activities - the Cyber Diplomacy Toolbox (Council of the European Union, 7 June 2017). The EU Cyber Diplomacy Toolbox includes confidence building measures, awareness raising on EU policies, EU cyber capacity building in third countries, negotiations, dialogues and demarches, as well as sanctions (Council of the European Union, 7 June 2017; Ivan, 2019, 5). This toolbox enables the EU to contribute to conflict prevention, to strengthen the rules based order in cyberspace, including the application of international law and the norms of responsible state behavior, but also to raise awareness among the public and decision makers. Finally, it includes restrictive measures in order to keep cyberspace functional and secure.

On 16 April 2018, the European Council adopted conclusions on malicious cyber activities which pointed out the importance of a global, open, free, stable and secure cyberspace and it mentioned also the threat of malicious activities in cyberspace (Council of the European Union, 16 April 2018). Later that year, on 18 October, the European Council requested from the member-states to develop their cyber capabilities in order to respond and deter attacks and threats through cyberspace (Council of the European Union, 18 October 2018). Likewise, on 12 April 2019, the High Representative declared the need of promoting an open, stable and secure cyberspace, the respect of human rights, freedom and rule of law, the urging of actors to stop using cyberspace for malicious activities in order to facilitate their actions and the need for strengthening national and international cooperation in order to safeguard cyberspace (Council of the European Union, 30 July 2020).

On 17 May 2019, the Council Decision 2019/796 established a framework for restrictive measures to deter and respond to threats through cyberspace (Official Journal of the European Union, 17 May 2019). Through this framework, the EU could impose targeted restrictive measures in order to deter and respond to attacks through cyberspace against the EU or its member-states. The cyber sanctions will be imposed to cyberattacks, who originate or are carried out from outside the EU, or use infrastructure outside the EU, or are carried out by persons or entities established or operating outside the EU, or are carried out with the support of person or entities operating outside the EU (Official Journal of the European Union, 17 May 2019, 2). Specifically, the above framework permits the EU to impose for the first time sanctions against persons or entities who are responsible for cyberattacks or malicious activities. The sanctions include among others, travel bans on persons travelling in and out of the EU, the prohibition of making funds and the freeze of the asset of persons or entities.

Both Council Decision 2019/797 and Council Regulation 2019/796 enlisted six types of cyberattacks which could lead to the imposition of sanctions. These attacks are against the critical infrastructure, the essential services, critical state functions, the storage or processing of classified information, the government emergency response teams, against the EU institutions and CSDP missions and operations (Pawlak & Biersteker, 2019, 32). These attacks should have a significant effect, including attempted cyberattacks which constitute an external threat to

the EU or its member-states according to the article 1 of the Council Decision. Moreover the Council Decision 2019/797 stresses that member-states shall adopt measures to prevent the entry into or transit through their territories of individuals who are responsible for cyberattacks or the support to attempt cyberattacks.

The EU sanctions regime is a sophisticated and complex set of measures which enables the member-states to choose between a number of possible instruments that can be used in combination with other important EU documents and legally binding decisions (Moret & Pawlak, 2017, 2). The EU's cyber sanctions regime could trigger other likeminded nations to respond in a similar manner and thereby bolster emergent norms or reinforce existing ones (Pawlak & Biersteker, 2019, 5).

Even though the inclusion of sanctions as a policy option in the EU Cyber Diplomacy Toolbox, points to the right direction, we have to bear in mind the following constraints. First, the EU seems to navigate in uncharted waters, since there is no previous example of cyber sanctions regimes. Second, and despite the progress in harmonizing national cyber policies and legislations, there is still lack of consensus within the Union regarding definitions and approaches to cybersecurity. In particular, member-states differentiate on what constitutes a cyberattack and what evidence is needed in order to attribute it. Third, not all member-states have the technical ability and necessary intelligence to attribute a cyberattack. Fourth, and in direct relation to the above point, the sanctions regime is not an entirely autonomous mechanism, since it is subject to judicial decision by the Court of Justice of EU. Actually, there are many cases of cyber sanctions which have been lost to the European Court of Justice due to insufficient proof by the EU Council. Fifth, cyber sanctions are not a panacea, but rather one more policy instrument. Thus, cyber sanctions must be combined with other policy instruments such as diplomacy, law enforcement, dialogue and cooperation with other likeminded countries and institutions. Six, cyber sanctions might influence state and state sponsored actors, but will have limited effect on non-state actors that are not dependent on their governments. Seven, the member-states through the procedure of public attribution, are taking the risk to expose important information and cyber capabilities. The cyber attackers can take advantage of these information for their activities to become more effective. The effectiveness of the cyberattacks can be from the identification of the vulnerabilities in member-states systems. A similar example is the cyberattacks of the WannaCry and NotPetya. The former was in 2017 and in October of the same year, information of this attack was published as evidence of attribution. The latter was more effective because it used all the evidence from the former positively. So the sharing of intelligence and cyber capabilities publicly should be in a more secure environment with limited access by non-competent actors (Pawlak & Biersteker, 2019, 59).

Getting back to the thorny issue of reliable attribution, the effectiveness of the cyber sanctions regime is highly depended on the ability to attribute a cyberattack. Apart from the technical parameters of identifying the source of an attack, attributing an attack to a specific actor, country, company or individual, is after all a political decision. Having constructed a theoretical understanding of sanctions and explored the cyber sanctions regime that has been established by the EU, we will now turn our attention to the empirical evidence of the recently imposed cyber sanctions.

## **5. Cyber sanctions imposed by the EU: A preliminary analysis**

On 30 July 2020, within the framework of the Common Foreign and Security Policy (CFSP), the European Council announced for the first time a regime of cyber sanctions (restrictive measures) with travel bans and asset freezes against six natural persons and three entities or bodies who were involved in a series of cyberattacks against the EU or its member-states, public and private sector (Official Journal of the European Union, 30 July 2020). These persons and entities or bodies are responsible for the following cyberattacks: WannaCry, NotPetya, Operation Cloud Hopper and the attempted cyberattack against the Organisation for the Prohibition of Chemical Weapons (OPCW). The sanctions targeted the following (Zlaikha, 2020, 2):

- Four members of Unit 74455 of Russia's military intelligence agency (GRU) for the NotPetya and the hacking of the Wi-Fi system for OPCW in Hague in April 2018. The Council also enlisted the GRU for the attacking the Ukrainian Electricity grid in 2015.
- Two Chinese citizens and the technological development company Huaying Haitai for their alleged involvement in an attack on information systems of companies in six continents and in the EU and service provider located in the EU such as Swedish Ericson, which was named the Operation Cloud Hopper.
- The Lazarus group, which consists of North Korean company Chosun Expo for the Wanna Cry ransomware attack, the collapse of the British NHS servers and millions of losses for the affected private sector (Official

Journal of the European Union, 30 July 2020). The WannaCry disrupted information systems around the world by targeting information systems with ransomware attacks and blocking access to data. It affected the information systems of companies inside the EU.

The attribution of cyberattacks to Chinese, Russian and North Korean entities and individuals is considered to be a bold move for the EU. It demonstrates clearly that the EU and its member-states have embraced the need to be more active than passive, when it comes to cybersecurity and cyberdefense. After the announcement of the cyber sanctions regime by the Council of the European Union, the US, the UK, Australia and Canada publicly supported this policy initiative. On the contrary, Russia and China criticized the EU for following a restrictive position instead of following a diplomatic tool like the dialogue. The imposition of cyber sanctions is similar to the concepts of 'active defence' or 'defending forward' both sponsored by the US (Zlaikha, 2020, p.3). According to the 2018 Department of Defence Cyber Strategy, the US will work with like-minded states in order to secure and deter the threats in cyberspace through the sharing of information, best practices, buttressing of attribution claims and public statements of support for responsive actions taken and joint action (US Department of Defense, 2018). Thus the US military can carry out activities in and out of cyberspace in order to collect information, to disrupt malicious cyber activities and respond to these activities below the level of armed conflict (Pawlak & Biersteker, 2019, 8).

On 11 September 2020, the Horizontal Working Party on Cyber Issues started the discussions for a second cyber sanctions package by the EU (Council of the European Union, 8 September, 2020). The Bundestag hack on 2015 triggered the EU for the second package of cyber sanctions. This cyberattack resulted in the exfiltration of 16GB of data of the German Parliament's information technology network. On October 2020, the European Council announced the second package of cyber sanctions against the head of GRU, Igor Kostyukov and the GRU officer Dmitriy Badin and the GRU Unit 26165 (also known as APT28). The European Council had already announced sanctions for Kostyukov with a travel ban and asset freeze in January 2019 for the Salisbury chemical attack on Sergei Skripal and his daughter and German authorities issued an arrest warrant for Badin in May 2020 for the Bundestag hack. The second cyber sanctions package referred to more EU travel restrictions on Badin, since they cannot arrest him because he is not on EU territory. On 5 October, the draft for the sanctions regime proposal was forwarded to the Working Party of Foreign Relations Counsellors and on 19 October the Committee of Permanent Representatives agreed to adopt the new sanctions regime. After this, the European Council announced a list of sanctions for Igor Kostyukov, head of the GRU and Dmitriy Badin, officer of the GRU and the GRU Unit 26165 (known as APT28) (Council of the European Union, 22 October, 2020). Only six out of twenty seven EU member states publicly endorsed the second cyber sanctions package.

Cyber sanctions on their own are not enough to prevent, deter and respond to malicious cyber activities. The success of cyber sanctions regimes depend also on a strong cooperation with the private sector and with non-EU member states. In the near future, the EU should adopt alternative tools instead of sanctions in order to respond to cyber malicious activities and also examine the option of not being constrained by the public attribution of its member-states.

## **6. Concluding remarks**

The EU, as any other global actor, faces enormous security challenges that relate to cyberspace. Over the last years, the EU has established a number of policies, strategies and institutions in order to counter the threats that derive from this domain. The Cyber Diplomacy Toolbox that was established in 2017, enables the EU to impose cyber sanctions. In 2020, the EU decided to make use of this policy tool and impose cyber sanctions for the first time. The imposition of these restrictive measures, highlighted once more the question of whether sanctions are truly effective or whether their utility is restricted in their symbolic nature.

Adding to that, the imposition of cyber sanctions by the EU exposed two more caveats. The first one relates to the unique and complex nature of cyberspace, where reliable attribution remains not only a technological challenge, but also or rather more, a political one. The second caveat and probably the most important one relates to the EU's political willingness to become a reliable (cyber)security provider.

Brussels have to choose between attributing cyberattacks and thereby jeopardising its relations with other countries, and not responding to cyber threats with cyber sanctions and thus choosing a more preferable policy option. So far the record is mixed. The EU imposed cyber sanctions, but lacked public support, since only six out of the twenty seven member-states publicly endorsed the second cyber sanctions package (Soesanto,



November, 2020). Till now, the imposition of cyber sanctions are against individuals, although threats and attacks are designed and ordered by states. As a result, the perpetrators could unleashed cyberattacks against the same targets or new targets in the near future.

This is a vicious circle. Cyber sanctions are an ineffective tool of response, due to technical reasons such as the existence of unclear evidence, false links and untraceable perpetrators. Nevertheless, the political value of the cyber sanctions regime lays in its symbolism. By imposing cyber sanctions, the EU demonstrates its willingness to respond and preserve its core values and its sovereignty (Erskine, 2020). It is safe to conclude that the cyber sanctions regime is one more option in the EU's toolbox. Whether cyber sanctions will prove to be a hammer or not, remains to be seen.

## **Acknowledgements**

This work has been partly supported by the University of Piraeus Research Center.

## **References**

- Baldwin A. David, (1985) *Economic Statecraft*, Princeton University Press, New Jersey.
- Council of the European Union, (16 April 2018), '*Council conclusions on malicious cyber activities – approval*', Brussels <https://data.consilium.europa.eu/doc/document/ST-7925-2018-INIT/en/pdf>
- Council of the European Union, (7 June 2017), *Draft Council Conclusions on a Framework for a Joint EU Diplomatic Response to Malicious Cyber Activities ("Cyber Diplomacy Toolbox")*, Brussels <https://data.consilium.europa.eu/doc/document/ST-9916-2017-INIT/en/pdf>
- Council of the European Union, (30 July 2020), '*Declaration by the High Representative Josep Borrell on behalf of the EU: European Union response to promote international security and stability in cyberspace*', Brussels <https://www.consilium.europa.eu/en/press/press-releases/2020/07/30/declaration-by-the-high-representative-josep-borrell-on-behalf-of-the-eu-european-union-response-to-promote-international-security-and-stability-in-cyberspace/>
- Council of the European Union, (18 October 2018), '*European Council meeting– Conclusions*', Brussels <https://www.consilium.europa.eu/en/press/press-releases/2018/10/18/20181018-european-council-conclusions/>
- Council of the European Union, (8 September, 2020), '*Horizontal Working Party on cyber issues*', Brussels <https://data.consilium.europa.eu/doc/document/CM-3442-2020-INIT/en/pdf>
- Council of the European Union, (22 October, 2020), '*Malicious cyberattacks: EU sanctions two individuals and one body over 2015 Bundestag hack*', Press office - General Secretariat of the Council, Brussels <https://www.consilium.europa.eu/en/press/press-releases/2020/10/22/malicious-cyber-attacks-eu-sanctions-two-individuals-and-one-body-over-2015-bundestag-hack/pdf>
- Erskine Sasha, (12 October 2020), '*The EU Tiptoes into Cyber Sanctions Regimes*', RUSI Library <https://rusi.org/commentary/eu-tiptoes-cyber-sanctions-regimes>
- European Union External Action, (3 August 2016), '*European Union sanctions*' [https://eeas.europa.eu/headquarters/headquarters-homepage/423/european-union-sanctions\\_en](https://eeas.europa.eu/headquarters/headquarters-homepage/423/european-union-sanctions_en)
- Ivan Paul, (18 March, 2019) '*Responding to cyberattacks: prospects for the EU Cyber Diplomacy Toolbox*', European Policy Centre <https://www.epc.eu/en/Publications/Responding-to-cyberattacks-EU-Cyber-Diplomacy-Toolbox~218414>
- Kapsokoli Eleni, (2020) '*EU Cybersecurity Governance: A Work in Progress*', in "Views on the Progress of CSDP", ESDC 1st Summer University Book, edited by Fotini Bellou and Daniel Fiott, European Security and Defence College, Doctoral School on CSDP, Publications Office of the European Union, Luxembourg [https://www.researchgate.net/publication/346274333\\_Eleni\\_Kapsokoli\\_EU\\_Cybersecurity\\_Governance\\_A\\_Work\\_in\\_Progress\\_in\\_VIEWS\\_ON\\_THE\\_PROGRESS\\_OF\\_CSDP\\_ESDC\\_1st\\_Summer\\_University\\_Book\\_edited\\_by\\_Fotini\\_Bellou\\_and\\_Daniel\\_Fiott\\_European\\_Security\\_and\\_Defence\\_C](https://www.researchgate.net/publication/346274333_Eleni_Kapsokoli_EU_Cybersecurity_Governance_A_Work_in_Progress_in_VIEWS_ON_THE_PROGRESS_OF_CSDP_ESDC_1st_Summer_University_Book_edited_by_Fotini_Bellou_and_Daniel_Fiott_European_Security_and_Defence_C)
- Liaropoulos Andrew, (July 2018), '*The Uses and Limits of Cyber Coercion*', Proceedings of the 17th European Conference on Cyber Warfare and Security, University of Oslo, 28-29 June 2018 [https://www.researchgate.net/publication/326265321\\_The\\_Uses\\_and\\_Limits\\_of\\_Cyber\\_Coercion\\_Proceedings\\_of\\_the\\_17th\\_European\\_Conference\\_on\\_Cyber\\_Warfare\\_and\\_Security\\_University\\_of\\_Oslo\\_28-29\\_June\\_2018](https://www.researchgate.net/publication/326265321_The_Uses_and_Limits_of_Cyber_Coercion_Proceedings_of_the_17th_European_Conference_on_Cyber_Warfare_and_Security_University_of_Oslo_28-29_June_2018)
- Moret Erica and Pawlak Patryk, (July 2017), '*The EU Cyber Diplomacy Toolbox: towards a cyber sanctions regime?*', European Union Institute for Security Studies <https://www.iss.europa.eu/sites/default/files/EUISSFiles/Brief%2024%20Cyber%20sanctions.pdf>
- Official Journal of the European Union, (17 May 2019), '*COUNCIL REGULATION (EU) 2019/796 of 17 May 2019 concerning restrictive measures against cyberattacks threatening the Union or its Member States*', Brussels <https://eur-lex.europa.eu/legal-content/GA/TXT/?uri=CELEX%3A32019R0796>
- Official Journal of the European Union, (30 July 2020), '*COUNCIL DECISION (CFSP) 2020/1127 amending Decision (CFSP) 2019/797 concerning restrictive measures against cyberattacks threatening the Union or its Member States*', Brussels <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32020D1127>

**Eleni Kapsokoli**

- Pawlak Patryk and Biersteker Thomas, (October, 2019) '*Guardian of the Galaxy: EU cyber sanctions and the norms in cyberspace*', Chaillot Paper/155 <https://www.iss.europa.eu/content/guardian-galaxy-eu-cyber-sanctions-and-norms-cyberspace>
- Soesanto Stefan, (20 November, 2020), '*Europe has no strategy on cyber sanctions*', Lawfare Institute in Cooperation with Brookings <https://www.lawfareblog.com/europe-has-no-strategy-cyber-sanctions>
- Taylor Brendan, (2010), '*Sanctions as Grand Strategy*', 1<sup>st</sup> Edition, Routledge.
- US Department of Defense, (2018), '*Summary 2018 Department of Defense Cyber Strategy*', USA [https://media.defense.gov/2018/Sep/18/2002041658/-1/-1/1/CYBER\\_STRATEGY\\_SUMMARY\\_FINAL.PDF](https://media.defense.gov/2018/Sep/18/2002041658/-1/-1/1/CYBER_STRATEGY_SUMMARY_FINAL.PDF)
- Zlaikha Vered, (24 August 2020), '*For the First Time, EU Sanctions in Response to Cyberattacks: Enhanced Deterrence Efforts by Western Countries?*', INSS Insight No. 1364 <https://www.inss.org.il/publication/european-sanctions-on-cyber-activity/>

# How the Civilian Sector in Sweden Perceive Threats From Offensive Cyberspace Operations

Joakim Kävrestad<sup>1</sup> and Gazmend Huskaj<sup>1, 2, 3</sup>

<sup>1</sup>School of Informatics, University of Skövde, Sweden

<sup>2</sup>Department of Military Studies, Swedish Defence University, Stockholm, Sweden

<sup>3</sup>Center for Asymmetric Threat and Terrorism Studies, Swedish Defence University, Stockholm, Sweden

[Joakim.kavrestad@his.se](mailto:Joakim.kavrestad@his.se)

[Gazmend.huskaj@fhs.se](mailto:Gazmend.huskaj@fhs.se)

DOI: 10.34190/EWS.21.106

**Abstract:** The presence of state-sponsored actors executing offensive cyberspace operations (OCO) has been made evident in recent years. The term offensive cyberspace operations encompass a range of different actions, including cyberespionage, disinformation campaigns, spread of malware and more. Based on an analysis of past events, it is evident that state-sponsored actors are causing harm to the civilian sector using OCO. However, the degree to which civilian organizations understand the threat from state-sponsored actors is currently unknown. This research seeks to provide new a better understanding of OCO and their impact on civilian organizations. To highlight this domain, the case of the threat actor Advanced Persistent Threat 1 (APT1) is presented, and its impact on three civilian organizations discussed. Semi-structured interviews were used to research how the threats from OCO and state-sponsored actors are perceived by civilian organizations. First, a computational literature review was used to get an overview of related work and establish question themes. Next, the question themes were used to develop questions for the interview guide, followed by separate interviews with five security professionals working in civilian organizations. The interviews were analysed using thematic coding and the identified themes summarized as the result of this research. The results show that all respondents are aware of the threat from OCO, but they perceive it in different ways. While all respondents acknowledge state-sponsored actors as a threat agent executing OCO, some respondent's argue that state-sponsored actors are actively seeking footholds in systems for future use while others state that the main goal of state-sponsored actors is to steal information. This suggests that the understanding of the threat imposed by OCO is limited, or at least inconsistent, among civilian security experts. As an interview study, the generalisability of this research is limited. However, it does demonstrate that the civilian society does not share a common view of the threat from state-sponsored actors and OCO. As such, it demonstrates a need for future research in this domain and can serve as a starting point for such projects.

**Keywords:** cybersecurity, state-sponsored, advanced persistent threat, civilian, offensive cyberspace operations

---

## 1. Introduction

The presence of state-sponsored actors performing offensive cyberspace operations (OCO) against civilian organisations is a fact (Osawa, 2017, Rowe, 2019). Being aware of, and understating your adversary is a crucial part of establishing defence capabilities (Beckett, 2017). However, the degree to which civilian organizations understand the threat from state-sponsored actors is currently unknown. This research seeks to provide new a better understanding of OCO and their impact on civilian organizations.

OCO are defined as a sequence of planned actions executed by an organized group of people with a defined purpose in and through hardware and software which are used to create, process, store, retrieve and disseminate information in different types of interconnected networks that build a large, global network, built and used by people (Huskaj and Wilson, 2020). The element "offensive" entails actions conducted by an organized group of people belonging to a rule-based nation-state attacking confidentiality, integrity, and availability of an adversary's information systems and related infrastructure. The purpose of frameworks for OCO "is not to facilitate the destruction of adversary military infrastructure, but rather to enable a military organization in rendering an adversary (both military and non-military) incapable to conduct an attack (both in cyberspace and in the physical domain)" (Huskaj and Iftimie, 2020). These insights are the results of ongoing and growing academic research on OCO (Huskaj, 2019, Iftimie, 2019).

Adversaries and various threat actors, however, are not bound by similar restrictions. They use OCO and related methods for political and financial gain, and also to control their societies. Methods for OCO include, but are not limited to, spear-phishing, social engineering, man-in-the-middle, and buffer overflows (Huskaj and Wilson, 2020). Therefore, even if an attack is labeled as ransomware, espionage, or wiper, the initial method to gain

access to a target’s information system and related infrastructure is always an offensive action using an offensive method. Rather than describing in detail one or two cases, various attacks are described, by noting the intent of the attack and the resulting impact.

In this paper, we demonstrate how OCO can impact civilian organizations by presenting a reviewing the case of “Advanced Persistent Threat 1” (APT1). We then perform a computational literature review with the goal of establishing what themes that are discussed under the umbrella of OCO targeting companies or organizations before we interview cybersecurity professionals from the civilian sector with the aim of addressing the research question of this research:

RQ: How do cybersecurity professionals in the civilian sector perceive the threat from OCO?

The results presented in this paper contributes to research with a better understanding of how OCO and state-sponsored actors are perceived by the cybersecurity professionals in the civilian society. The rest of this paper is structured as follows; the case of APT1 and the impact of OCO is presented in Section 2. The methodology for this research is described in Section 3 and the results are presented in Section 4. Section 5 presents the answer to the research question and the conclusions from this research.

## **2. The case – APT 1**

Advanced Persistent Threat 1 (APT1) was a threat actor stealing “intellectual property from English- speaking organizations” (Mandiant, 2013). According to the same report, “APT1 is believed to be the 2nd Bureau of the People’s Liberation Army (PLA) General staff Department’s (GSD) 3rd Department (总参三部二局), which is most commonly known by its Military Unit Cover Designator (MUCD) as unit 61398 (61398 部队)” (p. 3). Mandiant (2013) writes that “APT1 has systematically stolen hundreds of terabytes of data from at least 141 organizations, and has demonstrated the capability and intent to steal from dozens of organizations simultaneously” (p. 3). The longest time APT1 maintained access in a target’s network “was 1,764 days, or four years and ten months” (p.3). The targets included, but are not limited to, information technology, transportation, high-tech electronics, financial services, engineering services, satellites and telecommunications, energy, construction and manufacturing, aerospace, education, healthcare and, metals and mining (Mandiant, 2013).

The primary method of gaining access to a target organization’s network was using the offensive method of spear phishing. These e-mails were relevant to the recipient and prepared with one or several attachments, and/or a link. The attachments were disguised as zip-files, or PDF-files, or were actual zip-files that required a password to unzip them. The disguised zip and PDF-files were actually executables: even though the extension was zip or PDF, it included “119 spaces after ‘.pdf’ followed by ‘.exe” (Mandiant, 2013). Executing such a file would open a backdoor allowing a remote threat agent to conduct actions on the targets information systems. The other tactic was to use password-protected zip-files to bypass firewalls: a firewall cannot scan the contents of password-protected zip-files for malicious files.

The impact of offensive operations dubbed espionage operations are difficult to measure. It usually takes years before the impact is seen. In 2017, F-Secure, a cybersecurity company, reviewed the impact of the attacks on three companies (Hyvärinen, 2017): SolarWorld, Westinghouse Nuclear, and ATI Metals. Table 1 depicts the impact of the attacks: SolarWorld filed for bankruptcy, Westinghouse Nuclear declared bankrupt, and ATI Metals was trading at less than half of pre-attack levels. In 2020, an additional review of the companies’ status was done by the authors as noted in Table 1.

**Table 1:** The consequences of offensive operations depicted as espionage attacks

Who?	Status before attack	What was stolen?	Status as of 2017	Status as of 2020
SolarWorld	World-leader in solar technology	intellectual property, pricing information	Filed for bankruptcy May 2017	Reorganised in late 2017, and bankrupt again in March 2018
Westinghouse Nuclear	World-leader in nuclear reactor designs	Technical and design specifications	Declared bankrupt	Acquired by Brookfield Business Partners LP after 17 months of bankruptcy organization

Who?	Status before attack	What was stolen?	Status as of 2017	Status as of 2020
ATI Metals	World-leader in specialist metals	User credentials for every account on the IT estate	Trading at less than half of pre-attack levels	Trading at less than half of pre-attack levels

The President of the United States “issued a Memorandum to the Trade Representative stating inter alia that:” (p. 4).

*“China has implemented laws, policies, and practices and has taken actions related to intellectual property, innovation, and technology that may encourage or require the transfer of American technology and intellectual property to enterprises in China or that may otherwise negatively affect American economic interests. These laws, policies, practices, and actions may inhibit United States exports, deprive United States citizens of fair remuneration for their innovations, divert American jobs to workers in China, contribute to our trade deficit with China, and otherwise undermine American manufacturing, services, and innovation (USTR, 2018)”.*

The investigation shows that the Chinese-based offensive operations targeting these companies were not random, they were targeted when they had a business relationship or problem with china (USTR, 2018). Each company is discussed in more detail below.

### **3. Methodology**

As described in the previous sections, state-sponsored OCO do affect civil organization in different ways. The civilian society is, in this case, defined as non-military organizations in the public and private sectors. This research was conducted using semi-structured interviews as described by Robson and McCartan (2016) and preceded by a computational literature review (CLR) used to get an overview of related literature (Mortenson and Vidgen, 2016). The CLR was used to identify topics discussed under the domain of OCO targeting civilian organizations. The topics identified were used to derive themes for interviews with cybersecurity professionals from civilian organizations in Sweden and those themes formed the basis of the interview guide.

The interviews were held with five participants working as cybersecurity professionals in civilian organizations in Sweden and the research process can be summarized as follows:

- 1. A computational literature review as described by Mortenson and Vidgen (2016) was used to get an overview of the research field and establish question themes
- 2. The question themes were used to develop questions for the interview guide
- 3. Interviews were held separately with five participants
- 4. Interview recordings were transcribed
- 5. The transcribed interviews were analyzed using thematic coding as described by (Braun and Clarke, 2006)
- 6. The themes were summarized and used to answer the questions addressed in this paper

### **4. Results**

This research began with a computational Literature Review (CLR) which intended to outline what themes that were discussed in the domain of OCO targeting civilian organizations. The results from the CLR was used to form themes for the subsequent interviews and the CLR followed the methodology outlined by Mortenson and Vidgen (2016) and (Kunc et al., 2018). The Scopus databased was used with the following query:

*(((cyber\*) AND attack\*)) OR (((offensive) AND cyber\*) AND operation\*) OR (((computer) AND network) AND attack\*)) AND (((organisation\*) OR company) OR companies))*

The search resulted in 1511 hits. 1466 papers remained after removing papers without abstracts, authors and duplications. Using the CLR analysis procedure involves deciding on a number of topics to be established from the body of literature. The analysis relies on titles, keywords and abstracts of included papers and automatically outputs the themes that are most prominent based on the used words. Using between 10 and 100 themes is common and the number of themes is established by testing (Kunc et al., 2018). Eventually, 60 topics were created in this study resulting in 60 word-clouds that reflected the central words for each topic. An example is given in Figure 1, below.



**Figure 1:** Example topic word-cloud

Next, the 60 topics were manually clustered into 40 topics, as listed in Table 2 in alphabetical order.

**Table 2:** The characteristics of the 1466 articles, clustered into 40 topics

Topic	Topic name	Papers	Topic	Topic name	Papers
1	Attack on Water Systems	11	21	Industrial Control Systems	60
2	Attacks on companies & mail/web/networks	91	22	Information Systems	15
3	Attacks on DNS	25	23	Insider threat	38
4	Attacks on healthcare	27	24	Internet attacks	43
5	Attacks on IS in CNI and Nuclear Systems	84	25	Internet freedom	8
6	Attacks on information	60	26	IoT devices & attacks	49
7	Attacks on Information Systems	58	27	Malicious threat attacks	60
8	Attacks on network services	19	28	Modelling	21
9	Attacks on systems	81	29	Network attacks	97
10	Attacks on wireless/mobile network	15	30	Network detection & attacks	95
11	Business information	19	31	Phishing attacks	29
12	China-based attacks	22	32	Protection schemes against attacks	21
13	Cloud computing	30	33	Ransomware	20
14	Cybercriminal attacks	66	34	Smart grid	33
15	Cyber-/information security risk	45	35	Software attacks	23
16	Cyberspace attacks	18	36	Supply chain	26
17	Data cloud	19	37	Terrorist attacks	17
18	Data inspection	13	38	Threats to data	15
19	DDoS attacks	31	39	Virtualization attacks	17
20	Digital evidence	23	40	Web-based attacks	22

It is noteworthy that the top three topics cover network attacks (97 articles), network detection & attacks (95 articles) and attacks on companies & mail/web/networks (91 articles). The bottom three topics cover data inspection (13 articles), attacks on water systems (11 articles) and Internet freedom (8) articles. Furthermore, from a threat actor perspective, the topics cover hackers (11, 21), insider threats (23), and cyber-criminals (14). The topics identified were reviewed by the researchers and abstracted to the following central themes that were used as themes for the interviews:

- Perceived threat actors and threats from different actors
- Perceived direct and indirect threats from offensive cyberspace operations
- Attack vectors used by state-sponsored actors
- Technical defense, deterrence and monitoring measures

- **Strategical defense**

Five interviews were held separately with security professionals working in civilian organization. The interviews were transcribed and analyzed using thematic coding using the just presented central themes, as described by Braun and Clarke (2006). Inter-coder reliability was built into the coding process by letting one researcher code the majority of the interviews while another researcher coded some interviews and performed consistency checks on the other interviews, similar to Rose et al. (2016). The responses in each theme were then summarized and the summaries are presented at the end of this chapter as results of the interview process.

The respondents are kept anonymous but their background can be described as follows:

- R1: The respondent is now working as an information security coordinator at a Swedish agency. The respondent previously worked as CISO (Chief Information Security Officer) at another agency and was prior to that employed in the public sector in the service desk. She has about 7 years of experience in the security field.
- R2: The respondent's background is as a computer forensic examiner at a Swedish agency and at a private company. He has worked with digital forensics for about 8 years and worked as a developer for about 6 years before that.
- R3: The respondent works as a cyber-security consultant and works with security architecture.
- R4: The respondent has been working in IT for about ten years, and with security for eight of those years. He has been working as a forensic expert, but also with technical and strategic information security and risk management.
- R5: The respondent has been working with security for about 20 years. He has been working with everything from strategic security to technical security. He is currently working as a security consultant and has many customers in the critical infrastructure sector.

The remainder of this section will summarize the interview data that was gathered from each theme.

#### **4.1 Threat actors**

Discussing threat actors in general, all respondents discuss organized criminal organizations and state-sponsored actors as the currently most prominent threat actors. The underlying meaning in all interviews is that other threat actors are out there but they are not as capable as state-sponsored actors or organized criminal organizations and do not need to be the primary concern when establishing defense. The interviews suggest that all respondents are aware of state-sponsored actors as a threat agent.

#### **4.2 Threat from OCO**

The respondents paint different pictures when discussing the threat from state-sponsored actors. Two respondents describe that the purpose of OCO is for foreign states to get a foothold in systems to enable them to launch cyberattacks as part of armed or diplomatic conflicts. Two other respondents discuss that state-sponsored actors mainly want to steal intellectual property, while the fifth respondent is not at all specific. One respondent also states that foreign states use disinformation campaigns to influence other nations in, for instance, elections.

The respondents agree that one aspect that signifies foreign states as threat actors is that they have access to more time and resources than other threat actors. That makes them pose a unique threat to the organizations they target. However, foreign states are not threat agents for all organizations.

Another threat from OCO that was discussed during the interviews was the risk of being harmed as collateral damage. The respondents agree that the risk of being collateral damage is indeed large, especially if your organization cooperates with organizations or states that are high-value targets for state-sponsored actors. When asked about the risk of being harmed as collateral damage, one respondent replied: "Ask MAERSK". He explains that MAERSK suffered severely as a result of an attack against Ukraine that was supposedly attributed to Russia. Another respondent mentions attacks against critical infrastructure or cyber-critical infrastructure as attacks that would impact the own organization.

### **4.3 Attack vectors used by state-sponsored organizations**

An important part of any risk-based security work is understanding the attack vectors that threat actors may use. As such, the respondents were asked about what attack vectors state-sponsored organizations use. In general terms, the respondents agree that state-sponsored actors use the same attack vectors as other threat actors. There are, however, some attack vectors or modus, that are more commonly used by state-sponsored organizations than by others. The respondents discuss that these are attack vectors that are hard to defend against, but also expensive for an attacker to use. The attack vectors that were discussed during the interviews are described below.

- Long term Social engineering - The respondents discuss that a characteristic of state-sponsored actors is that they act with a long time span. One respondent describes that state-sponsored actors can have a time horizon of 20 years or more, enabling them to employ long term social engineering attacks where they, for instance, position insiders in targeted organizations for later use.
- Human intelligence - somewhat similar to social engineering, one respondent describes that state-sponsored actors employ a human intelligence-based approach to find employees in organizations that can be used as an attack vector. This includes actions such as intelligence gathering as preparation for spear-phishing or insider recruitment. One respondent describes that OCO may include identifying employees that are, for instance, disgruntled or in debt and leverage that information in order to recruit them as insiders.
- Zero-day exploits - Three respondents describe that the large amounts of resources available to state-sponsored organizations allow them to create and hold zero-day exploits. While the respondents' state that the use of zero-day exploits is common amongst several threat actors, they are expensive, high-value possessions. While criminal organizations will use zero-day exploits if they believe that they will gain enough from their attack, state-sponsored actors are more prone to excessive use of zero-day exploits to reach their objectives. One respondent describes that several zero-day exploits were used in an attack against a nuclear facility in Natanz (Iran) making that attack very expensive

To conclude what attack vectors the respondents perceived that state-sponsored actors use, they use all available attack vectors. What differs from other threat actors is that they are more motivated and better funded and can, therefore, use attack vectors that are more expensive and time-consuming.

### **4.4 Technical countermeasures**

The respondents do not think that there are any distinct technical measures that have to be implemented to mitigate the threat from OCO. However, two respondents stress the need for detection and logging mechanisms and state that detection is crucial in order to detect intrusions from state-sponsored actors. He describes that state-sponsored actors often maintain footholds in compromised systems for a long period of time and being able to detect an intruder can enable mitigation.

### **4.5 Strategical defense**

When discussing defense against OCO and state-sponsored actors in general, all respondents agree that it is hard because of the resources available to that specific threat actor. Two respondents specifically state that a key part of the defense is a risk-based structured security approach. This includes identifying key assets, understanding the threats against these assets and employ reasonable preventive measures. The data from the interviews make it clear that defending against state-sponsored actors is different, mainly because it is harder. Something that is very clear is that the defense must include human factors of information security and governance and include policies, strategies and awareness. One respondent specifically stated that the human element of security is crucial for defense against state-sponsored actors, he said that "If you make it impossible, or very hard, to attack the organization using the cyber domain, the attacker will try to find a way to attack the organization easier by using the human domain.

The data gathered from the interviews do not necessarily suggest that any new measures have to be taken, but the information security work has to be done. For instance, the interviews suggest that OCO often includes the use of zero-day exploits, which means that firewall, detection systems and likewise must be in place to enable detection. As another example, the interviews suggest that OCO includes insiders and espionage making it important to know whom you are working with, keeping control of your employees and such.



## **5. Discussion and conclusions**

The research question addressed in this study was “How do cybersecurity professionals in the civilian sector perceive the threat from OCO?” The short answer is that they are aware of the threat but perceive it in different ways, this is elaborated on below.

Based on the results of the interviews conducted in this research it is reasonable to conclude that cybersecurity professionals and the civil sector consider OCO to be a threat and state-sponsored actors to be a threat agent. This notion aligns well with the information presented in the background portion of this paper which exemplifies how the civilian sector has been affected by OCO.

It is, however, noticeable that the perception of the threat from state-sponsored actors differs to a large extent among the respondents. Some respondents describe that state-sponsored organizations want to establish a foothold in systems and use that in case of armed or diplomatic conflict while other respondents claim that state-sponsored actors are mainly looking to steal information. This could suggest that the understanding of state-sponsored actors is, after all, lacking or at least inconsistent among security experts. There is also a chance that the discrepancy in answers stem from the fact the respondents are sure to have different backgrounds from different sectors were the threat is perceived differently. The image of a threat actor that is not yet fully understood is strengthened by the fact that the respondents agree that defense against state-sponsored actors is hard. This insight could suggest that the community of security experts is not yet fully aware of what state-sponsored actors do and how.

Furthermore, and as shown in Table 2 presented in section 4, the topics depict “Attacks on [insert target]”. Offensive operations are (mostly) about organized groups of people conducting actions targeting information systems. The methods include, but are not limited to, social engineering, buffer overflows and man-in-the-middle attacks. Therefore, to understand offensive operations and related methods requires a high understanding of the underlying technology. In addition, the respondents have noted that cybersecurity is about “human aspects.” This fits well with the notion of security as a process rather than a product (Schneider, 2000). Further, the versatility of attack vectors used by state-sponsored actors emphasize the notion of cybersecurity as a socio-technical property. That notions is well aligned with previous research into cybersecurity (Malatji et al., 2019). The respondents highlight that state-sponsored actors are more prone to using long-terms social engineering and human intelligence attack vectors, suggesting that the “human aspect” of security is even more important when considering the threat from OCO and state sponsored actors as compared to other threat agents.

To the best of our knowledge, little or no prior research maps how cybersecurity professionals in the civilian sector perceived the threat from OCO and their potential impact. The APT1-case presents the impact of OCO. Understanding how public and private organizations in civil society perceive OCO is imperative to future strategic information security practices. As such, this is a step towards a better understanding of how civilian organizations are affected by OCO carried out by state and state-sponsored actors. The impact is twofold, the interviews held in this study do show that OCO is a powerful threat that civilian organizations must consider in the information security practice, and those civilian organizations are worried by this threat. Second, the interviews emphasize the notion that a large and important part of information security work takes place in the human domain, technical defense alone is not enough.

In terms of validity, it should be noted that the intention of this research is to provide an initial outlook of how OCO and state-sponsored actors are perceived by civil society. Using interviews was selected because it allows for a discussion around the subject area and can provide an in-depth understanding of how the respondent perceived the area. On the other hand, it reflects the opinions of the respondents and the results can not necessarily be interpreted as general for all information society professionals globally or even in Sweden. another concern has to do with the nature of the subject area, not many security professionals are comfortable talking about the specifics about how their organization perceive or handle important threat actors, making the replies general in nature. to maximize the respondent’s ability to speak freely and ensuring that they were not put at risk in any way, they were guaranteed anonymity as a result of the ethical guidelines proposed by Schrittwieser et al. (2013).

The research presented in this paper suggests that OCO is a threat that the civilian sector is aware of. However, there is a need for a better understanding of how this particular threat is perceived and handled by the entire body of a civilian organization. One apparent direction for future work is to perform a large-scale survey study to research the same topics addressed in this paper in a bigger population. This research does also suggest that the civilian sector, as a whole, does not have a unanimous perception of how they can be threatened by state-sponsored organizations making the need for more research into how the civil society is affected by OCO apparent.

## References

- Beckett, P. 2017. Data And Ip Are The New Nuclear: Facing Up To State-Sponsored Threats. *Network Security*, 2017, 17-19.
- Braun, V. & Clarke, V. 2006. Using Thematic Analysis In Psychology. *Qualitative Research In Psychology*, 3, 77-101.
- Huskaj, G. The Current State Of Research In Offensive Cyberspace Operations. 18th European Conference On Cyber Warfare And Security (Eccws 2019), 4-5 July 2019, Coimbra, Portugal, 2019. Academic Conferences And Publishing International Limited, 660-667.
- Huskaj, G. & Iftimie, I. A. Toward An Ambidextrous Framework For Offensive Cyberspace Operations: A Theory, Policy And Practice Perspective. International Conference On Cyber Warfare And Security, 2020. Academic Conferences International Limited, 243-Xv.
- Huskaj, G. & Wilson, R. L. An Anticipatory Ethical Analysis Of Offensive Cyberspace Operations. International Conference On Cyber Warfare And Security, 2020. 512-520.
- Hyvärinen, N. 2017. Apt1 – What Happened Next? Available From: <https://blog.f-secure.com/apt1-what-happened-next/>.
- Iftimie, I. Cyber Sanctions: The Embargo Of Flagged Data In A Geo-Cultural Internet. European Conference On Cyber Warfare And Security, 2019. Academic Conferences International Limited, 668-Xiv.
- Kunc, M., Mortenson, M. J. & Vidgen, R. 2018. A Computational Literature Review Of The Field Of System Dynamics From 1974 To 2017. *Journal Of Simulation*, 12, 115-127.
- Malatji, M., Von Solms, S. & Marnewick, A. 2019. Socio-Technical Systems Cybersecurity Framework. *Information & Computer Security*.
- Mandiant 2013. Exposing One Of China's Cyber Espionage Units.
- Mortenson, M. J. & Vidgen, R. 2016. A Computational Literature Review Of The Technology Acceptance Model. *International Journal Of Information Management*, 36, 1248-1259.
- Osawa, J. 2017. The Escalation Of State Sponsored Cyberattack And National Cyber Security Affairs: Is Strategic Cyber Deterrence The Key To Solving The Problem? *Asia-Pacific Review*, 24, 113-131.
- Robson, C. & McCartan, K. 2016. *Real World Research*, John Wiley & Sons.
- Rose, J., Jones, M. & Furneaux, B. 2016. An Integrated Model Of Innovation Drivers For Smaller Software Firms. *Information & Management*, 53, 307-323.
- Rowe, B. I. 2019. Transnational State-Sponsored Cyber Economic Espionage: A Legal Quagmire. *Security Journal*, 1-20.
- Schneier, B. 2000. The Process Of Security. *Information Security*, 3, 32.
- Schrittwieser, S., Mulazzani, M. & Weippl, E. Ethics In Security Research Which Lines Should Not Be Crossed? Security And Privacy Workshops (Spw), 2013 Ieee, 2013. Ieee, 1-4.
- Ustr 2018. Findings Of The Investigation Into China's Acts, Policies, And Practices Related To Technology Transfer, Intellectual Property, And Innovation Under Section 301 Of The Trade Act Of 1974.

# Aviation Sector Computer Security Incident Response Teams: Guidelines and Best Practice

Faith Lekota and Marijke Coetzee  
University of Johannesburg, South Africa

[nombu30@gmail.com](mailto:nombu30@gmail.com)

[marijkec@uj.ac.za](mailto:marijkec@uj.ac.za)

DOI: 10.34190/EWS.21.028

**Abstract:** The digitisation of the aviation sector provides benefits for passengers and consumers while at the same time introducing complexity, as the integration of legacy systems with new technologies is not straightforward. In this process, aviation systems vulnerabilities are making the industry more susceptible to attacks from cybercriminals. Cybercriminals exploit system vulnerabilities to compromise aviation systems, leading to significant disruptions and compromise of air transport safety. The emergence of cyber-attacks within the aviation industry has led to establishing the aviation sector Computer Security Incident Response Teams (CSIRTs). There is a general recognition that it is essential to build cyber resilience into aviation systems by ensuring better management of cyber-attacks and a willingness to share information on challenges and solutions. Unfortunately, the aviation sector CSIRTs are not commonly found across the world. In this regard, both the European Union (E.U.) and the United States of America (USA) are taking the lead by providing effective cybersecurity incident response services to their constituencies. The Aviation Information Sharing and Analysis Center (A-ISAC) was established in the United States of America (USA), while Europe is pursuing its initiative via the European Centre for Cyber Security in Aviation (ECCSA). This paper aims to analyse established aviation CSIRTs both in the European Union and the United States of America. The critical aspects such as team composition, services provided to constituencies, dissemination of information using secure and effective methods, collaboration with other CSIRTs, including best practice normative standards and legal compliance. The paper further outlines challenges in the management of aviation CSIRTs. Guidelines and lessons learned from globally established aviation sector CSIRTs are gleaned due to the review to provide guidance when implementing an aviation sector CSIRT.

**Keywords:** aviation, CSIRTs, incident response, information sharing, collaboration, secure information dissemination

---

## 1. Introduction

The emergence of cyber threats and attacks globally has led to the establishment of cybersecurity incident response teams to manage cyber-attacks better. Various organisations have taken the initiative to manage cyber-attacks by developing incident response capabilities formally to manage the attacks and willingness to share information. Hámornik et al. (2017) allude that cybersecurity is becoming a critical threat globally, and organisations have set up specialised cyber-defence forces. Such arrangements for cyber defence groups is a Computer Emergency Response Team (CERT) or a Computer Security Incident Response Team (CSIRT) (Hámornik & Krasznay, 2017).

The domain-specific cyber-attacks within specific aviation groups has led to the establishment of CSIRTs to manage better cyber-attacks and a need to share information. While the aviation cybersecurity incident response teams are not typical and established globally, Europe is pursuing its initiative via the European Centre for Cyber Security in Aviation (ECCSA) (EASA, 2017a). Airbus has joined forces with other stakeholders, such as SITA (SITA, 2021), to launch a new Cyber Security Aviation Security Operations Center (SOC) (Airbus, 2017). Furthermore, industry stakeholders such as Boeing established a protected forum in which industry and government can exchange information about emerging information security cyber threats to the aviation industry. Within the USA, the Aviation Information Sharing and Analysis Center (A-ISAC) was established as a specialised forum to provide cybersecurity incident response services and manage risks to the aviation critical infrastructure (A-ISAC, 2020; Rechner et al., 2012).

The paper is structured as follows: Section 2 discusses aviation system cyber resilience to ensure aviation systems can recover after the attack. Section 3 discusses the established aviation sector CSIRTs within the European Union (E.U.) and the United States of America. A comparative study between established CSIRTs in both countries is discussed in section 4, outlining CSIRT model types, services they provide, and best practice standards. Section 5 discusses aviation CSIRTs challenges as experienced by other sector-based CERTs/CSIRTs. The last section 6 discusses issues to be considered at a regional level. Finally, the paper concludes.

## 2. Aviation ICT systems cyber resilience

This section discusses issues related to cyber-attacks that can disrupt systems, affect operations, cause the infrastructure to deteriorate, and the system to fail after the incident has occurred (Wang et al., 2020). Cyber resilience is the evaluation of the system before, during, and after a system encounters a threat due to unknown and unexpected threats from cyberspace activities (Eurocontrol, 2012; Lykou et al., 2019). As stated in the Federal Aviation Authority (FAA) (2018) report, cyber resilience remains a significant challenge for both the U.S. and Europe (FAA, 2018). One of the significant discussions among aviation stakeholders is how to sustain aviation ICT's efficient operability and resilience after a cyber-attack. The following section discusses the emergence of aviation sector Computer Emergency Response Teams (CERTs), Computer Security Incident Response Teams (CSIRTs) and Aviation Information Sharing Centers (A-ISACs) in the European Union (E.U.) and the United States of America (USA) to manage system resilience.

## 3. State of the aviation sector in Europe and the USA

First, the section discusses airspace management in the E.U. and U.S. regions. The following section discusses the established the E.U. aviation computer emergency response teams (CERT) and the U.S. Aviation Information Sharing Centers (A-ISAC) to manage system resilience and cyber-attacks. The discussion further outlines the services they provide and the communication model between CERT and A-ISAC.

### 3.1 Airspace management

Airspace management is a crucial aspect of the air industry to ensure passengers and other stakeholders' safety and security. Globally, organisations manage the airspace at regional or country levels. Different states adopt different airspace management, security models and assume responsibilities following the principles laid down by the International Civil Aviation Organisation (ICAO) (CANSO, 2016). The models are either centralised or decentralised. In the *centralised model*, security activities are primarily the state's responsibility. In the *decentralised model*, the airport authorities could provide the main security activities under the relevant authority's supervision, commonly the Civil Aviation Authority (CAA) (Ford et al., 2020).

The section begins with discussing the management of the airspace and the models adopted within Europe and the USA. The discussion includes outlining the airspace's size, the number of airports within each region, the average number of flights, and the staff number.

With emerging aviation cyber-attacks, both regions have established cybersecurity incident response teams at various levels to manage cyber incidents and maintain critical infrastructure resilience. A significant difference is that Europe consists of many individual sovereign states and has adopted a centralised and decentralised model to manage the European airspace and aviation security. As one sovereign state, the USA adopted the centralised model to manage airspace and aviation security. The adoption of different models is an indication that the aviation ecosystem is vast and complex to ensure that the entire environment is secure from imminent cyber-attacks.

Table 1 outlines the year 2018 comparative critical aspects between the E.U. and USA airspace management.

**Table 1:** Europe and USA critical Air Traffic Management (ATM) system

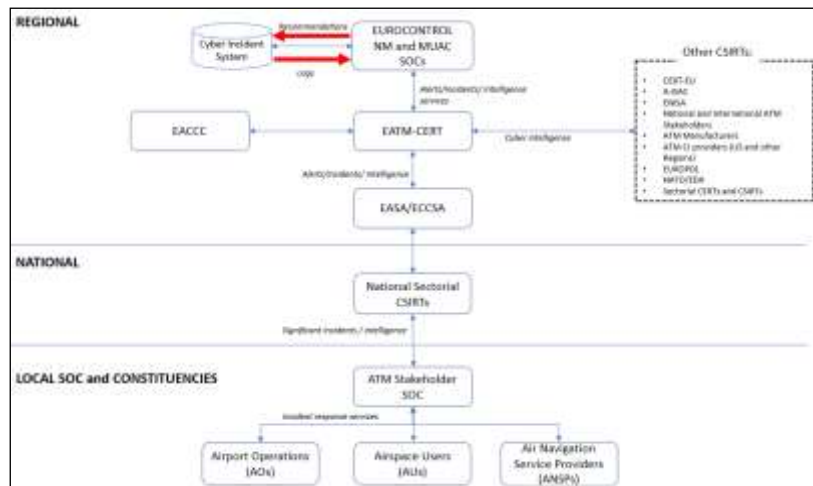
	Europe	USA	Comparison Europe vs USA
Million per square km	11,5	10,4	Europe airspace is 1,1 > bigger than the USA
Service Providers	37	1	37 ANSPs in Europe versus 1 ANSP in the USA
Stand alone approach control facilities	16	26	Less than 10 < standalone approaches than the USA
En-route facilities	62	20	More than 42 > en-route facilities in Europe
Airports with ATC services	406	517	USA has 111 > airports with ATCs
Average daily flights	28 475	41 874	USA has more 13 999 > average daily flights than Europe
Share of general aviation (IFR)	4%	19%	More IFR in USA < 15% than Europe
Total staff	55 130	31 647	More staff in Europe < 23 483
ATCOs in OPS	17 794	12 170	More ATCOs in Europe < 5624

Table 1 shows that the European airspace is more complex than the USA airspace. The USA has an average of 13 999 more daily air flights and has a more significant number of airports than Europe. We can deduce that the USA airspace could be congested and yet smaller than the European airspace from the data provided. However, a single Air Navigation Service Provider (ANSP) in the USA manages the airspace compared to Europe. The following section discusses the established European Union (E.U.) aviation sector CERT.

### 3.2 European Union aviation sector

The previous sections highlighted the E.U. complex airspace management and the challenges of cyber resilience. This section discusses the establishment of the E.U. cybersecurity incident response teams, followed by the incident response teams' services. Lastly, the following section discusses the information-sharing model among aviation stakeholders.

While, EUROCONTROL, dedicated to supporting European aviation (Eurocontrol, 2020), established the European Air Traffic Management Computer Emergency Response Team (EATM-CERT) in 2017 to manage cybersecurity incidents. Figure 2 outlines the EATM-CERT structural composition and resembles the author's interpretation from source documents (EASA, 2017c, 2020b; Eurocontrol, 2020).



**Figure 2:** Constitution of E.U. aviation sector CSIRT adapted from EUROCONTROL (EUROCONTROL, 2020a)

#### 3.2.1 Regional

The European Air Traffic Management Computer Emergency Response Team (EATM-CERT) serves as a coordination hub for major aviation Security Operations Centers (SOCs) at a regional level (Eurocontrol, 2020; Patrick Mana & Friligkos, 2019). As depicted in Figure 2, the EUROCONTROL cyber incident system is a central database containing cybersecurity incident data for future access and trend analysis. In 2017, the European Union Aviation Safety Agency (EASA), through a Memorandum of Understanding (MoU) with CERT-EU, established the European Centre for Cyber Security in Aviation (ECCSA). The ECCSA coordinates cyber intelligence information with national aviation CSIRTs (EASA, 2017a, 2017b, 2020b; P Mana, 2019).

#### 3.2.2 National

Since the European Union comprises many different states, it is difficult to have one agency to provide security for all different aviation stakeholders (Ford et al., 2020). Thus, Europe adopted a combination of centralised and decentralised models to manage airspace and security. Germany is an example of such a region where aviation security management is centralised, managing both the airspace and aviation security. However, in other European regions, the security model's adoption is decentralised (EASA, 2017a; P Mana, 2019). With the decentralised model, local Security Operations Centers (SOC) could be established.

#### 3.2.3 Local

The *local* level depicted in Figure 2 represents local ATM SOC that provides cybersecurity incident services to local constituencies such as Airport Operators (AOs), Airspace Users (AUs), and Air Navigation Service Providers (ANSPs) (Mana, 2019).

#### 3.2.4 E.U. Aviation CSIRT services

The EATM-CERT provides incident response support services to the Network Manager (NM), the cross-border Maastricht Upper Area Control Center (MUAC), and pan-European aviation stakeholders such as Air Navigation Service Providers (ANSPs) and Airport Operators (AOs) (Eurocontrol, 2020). The team further provides proactive

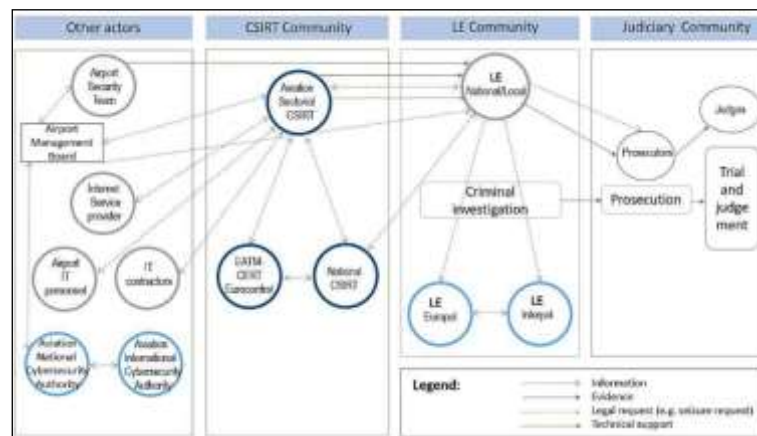
and reactive cyber-security monitoring services, supports national sectoral CERTs, and coordinates cybersecurity alerts and incidents (EUROCONTROL, 2020b; Patrick Mana & Friligkos, 2019).

Overall, the role of national sectorial CSIRTs is to collaborate with local Air Traffic Management (ATM) Security Operations Centers (SOC) on issues of significant cybersecurity incidents and cyber intelligence. At the local level, the SOC's primary function is to detect, analyse, respond, and report on cybersecurity incidents from aviation stakeholders.

The following section deliberates on the cybersecurity information sharing landscape across several incident response structures.

### 3.2.5 E.U. aviation sector CSIRT cybersecurity information sharing landscape

Cyber-related information sharing is critical between aviation CSIRTs and other established sectorial CSIRTs within the European Union. This section shows an orchestrated information flow between different actors within the aviation ecosystem (Anderson et al., 2021). As the central hub, the Aviation Sectorial CSIRT receives incident notifications from other aviation stakeholders, EATM-Cert, and National CSIRT. The cyber intelligence information shared with the Law Enforcement (L.E.) agencies signifies the management of cyber-attacks holistically. Possible roles and responsibilities are clearly defined, with a synergy between aviation CSIRTs, Law Enforcement (L.E.) and Judiciary communities as a concerted effort to fight cybercrime.



**Figure 3:** Graphical representation of information sharing and overview interaction (Anderson et al., 2021)

In summary, this section outlines the importance of collaboration and partnerships with Law Enforcement as paramount for effective cybersecurity management. Aviation stakeholders use the secure information-sharing platform to exchange domain-relevant cybersecurity information, vulnerabilities, and cybersecurity incident events (EASA, 2017a, 2017b, 2020b; P Mana, 2019). The following section discusses the developments of the U.S. aviation Information Security Analysis Center (A-ISAC).

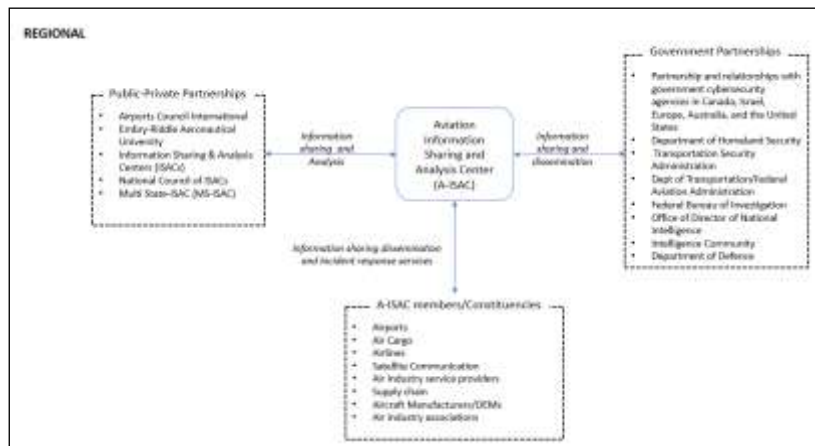
## 3.3 Aviation CSIRTs in the United States of America (USA)

Following the discussion about progress made in Europe, it is prudent to discuss such an initiative in the USA. First, this section introduces the established Aviation-Information Sharing Analysis Center (A-ISAC) in the USA. Next will be an overview of the structural formation of the A-ISAC and a discussion on collaboration between established incident response organisations. Following will be a discussion on the services that the team provide. Lastly is the discussion about the cybersecurity information sharing landscape between different functions.

### 3.3.1 USA Aviation-Information Sharing Analysis Center (A-ISAC)

The 2019 air traffic flow data show an increase in air traffic volume across the USA airspace (FAA, 2019). Although 2020 stats show a drop in the air traffic flow due to the global pandemic, the statistics still indicate that the number of aircraft across the USA airspace is high (CANSO, 2020). The Aviation Information Sharing and Analysis Centre (A-ISAC) was formally established in 2014 to maintain system resilience and security (A-ISAC, 2020; Francy, 2015).

First, the section discusses the established and composition of the A-ISAC and incident response services. Lastly, the section discusses the information-sharing landscape. Figure 4 illustrates the composition of A-ISAC and its collaborative model with other sectors. The diagram resembles the authors use of relevant source documents to design the structure. Dotted blocks in the diagram identify sectors that collaborate and share information with the A-ISAC.



**Figure 4:** Aviation Information Sharing Center (A-ISAC) adapted from A-ISAC organisation (Francy, 2015)

### 3.3.2 Regional

Figure 4 illustrates the established U.S. Aviation Information Sharing and Analysis Center (A-ISAC) at the regional level. The A-ISAC uses a centralised cybersecurity collaborative framework and a partnership model to share information between the government, public and private sectors, and aviation constituencies.

### 3.3.3 Public and private sector partnerships

The block on the left of Figure 4 resembles the public and private partners collaborating and sharing information with the Aviation Information Sharing and Analysis Center (A-ISAC) (Aviation ISAC, 2018; Francy, 2015).

### 3.3.4 Government partnerships

To the right of Figure 4 is the collaboration of A-ISAC with U.S. government entities. The A-ISAC formed partnerships with various entities and collaborations with international and national bodies (Francy, 2015).

### 3.3.5 A-ISAC members/constituencies

Figure 4 depicts A-ISAC members/Constituencies at the bottom. The A-ISAC offers incident response services to Airports, Air Cargo, Airlines, Satellite Communication, Air industry service providers, Supply chain, Aircraft Manufacturers/OEMs, and Air industry associations (Francy, 2015).

### 3.3.6 A-ISAC services to constituencies

Overall, the Aviation Information Sharing and Analysis Center (A-ISAC) provides a secure trust network, especially on sharing cyber intelligence. The centre conducts research and information analysis and validates the accuracy and severity of threats to its constituencies (A-ISAC, 2020). Furthermore, the A-ISAC offers industry best practices and cyber awareness training to aviation stakeholders (A-ISAC, 2020). The following section deliberates on the A-ISAC cybersecurity information sharing landscape with other sectors.

### 3.3.7 A-ISAC cybersecurity information sharing landscape

Figure 5 highlights the Aviation Information Sharing and Analysis Centre (A-ISAC) between multiple stakeholders. Colour icons are used in the figure and represent sector-specific stakeholders, where the dotted line indicates peer-to-peer information sharing. The solid line illustrates information sharing among stakeholders. The trust icon illustrated in the diagram resembles a trust framework and signed bilateral agreements between parties. The double-sided arrow indicates that the flow of information is a two-way process, and the one-sided arrow indicates the flow of information as one-way.







## 5. Challenges

This section discusses challenges experienced by globally established CSIRTs which are not unique to aviation CSIRTs. Drawing on an extensive range of sources, established authors have set out to discuss CSIRTs' challenges in managing their constituencies. Table 2 summarises the challenges experienced by aviation CSIRTs.

**Table 2:** Cybersecurity Incident response challenges

Identified challenges	Description
<b>Human factors</b>	There is a shortage of cybersecurity incident response skills (Hámornik & Krasznay, 2017).
<b>Fragmented technologies</b>	The Aviation industry is global and interconnected but remains fragmented (International, 2020).
<b>Information sharing</b>	Teams are unwilling to share information about identified vulnerabilities and reluctant to expose their constituency vulnerabilities (Bradshaw, 2015).
<b>Trust and confidentiality</b>	Lack of trust since organisations may be worried about reputational damage and publicising a cyber-attack (Goodwin et al., 2015).
<b>Legislative framework</b>	International and national laws impede the ability of CSIRTs to share data (Killcrece, 2005).
<b>Lack of secure communication channels</b>	Lack of secure communication channels impacts providing confidential information to the constituency and affects trust and confidentiality (Skierka et al., 2015).

Conclusions can be drawn from table 2 that there is a need to manage incident response challenges through risk mitigation action plans. The following section discusses a comparative analysis of the established aviation sector cybersecurity incident response teams both in Europe and the U.S.

## 6. A comparative study between E.U. and USA aviation CSIRTs

Briefly, the European Union and the United States have progressed well with implementing cybersecurity incident response teams to address aviation cyber-attacks and manage system resilience. The E.U. and U.S. regions have established aviation sector-based cybersecurity incident response teams that collaborate and share information with international partners. Furthermore, the U.S. and the USA have adopted the General Data Protection Regulation (GDPR) and a common trust framework to enable harmonisation and exchange of sensitive data information across all regions (FAA, 2018; Majka & Wacnik, 2019). Sensitive information is shared and disseminated through secure communication platforms. There are evident similarities between the U.S. and the U.K. regarding collaboration and information sharing among sectors.

Both organisations provide comparable cybersecurity incident response services to their constituencies and have adopted best practice information security standards and frameworks as a form of effective service delivery model. Similarly, aviation sector CSIRTs and A-ISACs experience cybersecurity incident response challenges experienced by other sector-based CERTs/CSIRTs. Lastly, the following section discusses issues to consider for establishing an aviation sector CSIRTs at the regional level.

## 7. Proposed generic aviation CSIRT at the regional level

Given the discussions about the established aviation cybersecurity incident response teams in Europe and the USA, there is an exciting opportunity to advance such frameworks at the regional level. The paper proposes a generic framework applicable to any region that does not have such an incident response framework. For example, a few countries around Australia, including Arica as a region, could not have a well-structured and defined aviation incident response framework. Figure 6 illustrates a proposed aviation sector CSIRTs structure at a regional level.

Figure 6 depicts regional aviation CSIRT that collaborates with other sector CERTS or CSIRTs—for example, the European CERT-EU and EATM-CERT and the regional crisis centre. The regional aviation CSIRT collaborates and share cyber intelligence information with relevant parties through a signed Memorandum of Understanding (MoU) and utilising a Trust framework. Furthermore, the regional aviation CSIRT acts as a central hub for national aviation CSIRTs and uses a threat analysis repository to store cyber intelligence data for further analysis and information sharing. National CSIRTs connect and share cybersecurity information at a peer-to-peer (P2P) level, collaborate with other sector-based national CERTs/CSIRTs, and share information with local aviation SOCs. The local aviation SOCs provide cybersecurity incident response services to the respective aviation stakeholders and further collaborates with other sector-based local CERTs/CSIRTs.

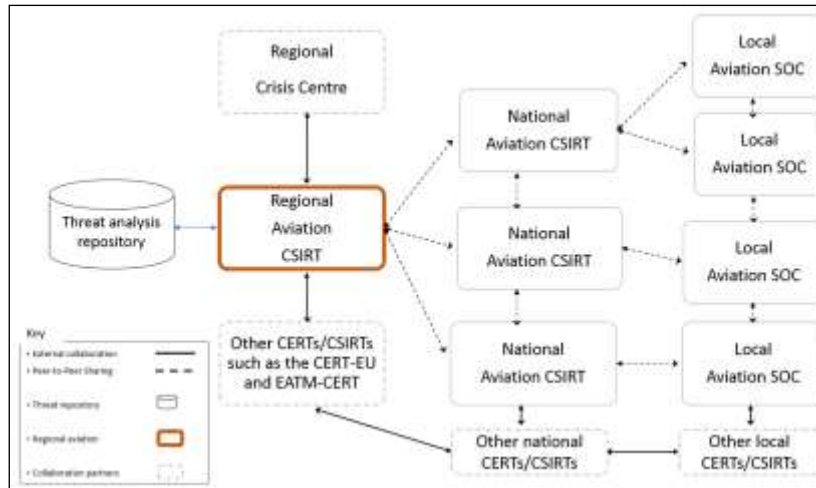


Figure 6: Proposed regional aviation sector CSIRT structural composition

To further elaborate on the proposed regional aviation sector CSIRT, table 3 defines an R-Responsible, A-Accountable, C-Consulted, I-Informed (RACI) matrix, a responsibility assignment chart that maps out every task assigned to key actors. of key actors. The following paragraph describes roles that stakeholders might play in the formation, execution and management of cybersecurity incident response:

- *Responsible*: Dedicated structures take full responsibilities for tasks within their domains;
- *Accountable*: Authorities and stakeholders are accountable for the completion, approval and signing-off of the task;
- *Consulted*: Authorities and stakeholders are consulted before a task is completed and signed-off; and
- *Informed*: The authorities and stakeholders are informed about decisions undertaken and consult with authorities once the decision is undertaken (CIO, 2021).

Table 3: RACI MATRIX

Tasks	Regional	National	Local
Develop regional aviation CSIRT strategy	R / A	C / I	C / I
Collaborate with the regional Crisis Centre and other sector-based bodies	R / A	C / I	C / I
Establish a regional crisis centre that will collaborate with the CSIRTs.	R / A	C / I	C / I
Establish national aviation CSIRTs and Security Operations Centers (SOCs)	C / I	R / A	R / A
Develop national and local aviation CSIRT strategy, and execute plans	C / I	R / A	R / A
Formulate a certification body for the aviation sector CSIRTs	R / A	C / I	C / I
Develop accreditation standards and training of aviation sector CSIRTs	R / A	C / I	C / I
Formulate standardised certification for aviation sector CSIRTs	R / A	C / I	C / I
Develop a common Legislative framework for sharing sensitive information	R / A	C / I	C / I
Implement a central database for the region as a central repository	R / A	C / I	C / I
Establish a regional crisis centre that will collaborate with the CSIRTs.	R / A	C / I	C / I
Provide aviation cybersecurity incident response services	R / A	R / A	R / A
Share cybersecurity incident and threat intelligence information	R / A	R / A	R / A
Adopt best practice standards and frameworks	R / A	R / A	R / A

The following session concludes the analyses of established aviation sector CERTs/CSIRTs and the best solution to deploy to regional aviation CSIRTs.

## 8. Conclusion

The following conclusion can be drawn from the paper that both E.U. and USA have been instrumental in establishing the aviation sector CSIRTs. The comparative analysis indicated the composition, essential services they provide, best practice frameworks, and standards. There is a harmonised collaborative and partnership approach among countries at an international level regarding cyber incident information sharing. The challenges faced by the aviation sector CSIRT require an effective cybersecurity strategy to mitigate the risk and maintain system resilience. In summary, trust and confidentiality are the cornerstones of establishing an effective aviation sector CSIRT. More creative solutions to building trust, information sharing, legislation, and best practice standards are essential guiding principles for implementing aviation cybersecurity incident response teams.

Suitable lessons are drawn from the European centralised and decentralised model to apply to any region without such a framework with clearly defined roles and responsibilities. Thus, the framework could serve as a guideline for managing incident responses in aviation within a region such as Africa.

## References

- A-ISAC. (2020). *Aviation Information Sharing & Analysis Center*. A-ISAC. <https://www.a-isac.com/>
- Airbus. (2017). *Cybersecurity Aviation SOC. SITA*, 1–16.
- Anderson, P., Bouza, B., Karkala, S., Kourtis, G., Michota, A., Patrascu, C., Portesi, S., Stupka, V., & Van Impe, K. (2021). *Aspects of cooperation between CSIRTS and Law Enforcement Handbook, Document for trainers* (Issue January). ENISA. <https://doi.org/10.2824/71834>
- Aviation ISAC. (2018). *MEDIA STATEMENT Aviation Industry Affirms the Safety of Commercial Aviation Aviation Information Sharing and Analysis Center sets the Record Straight on Safety and Security of Aircraft*. [www.a-isac.com](http://www.a-isac.com)
- Bradshaw, S. (2015). Best Practice Forum on Establishing and Supporting Computer Security Incident Response Teams (CSIRT) for Internet Security. *SSRN Electronic Journal*, 23. <https://doi.org/10.2139/ssrn.2700899>
- CANSO. (2016). *CANSO Charter and Articles of Association civil air navigation services organisation Edition 17-Approved at Vancouver AGM*. [https://www.canso.org/system/files/CANSO Articles of Association - Edition 17 - 2016 0.pdf](https://www.canso.org/system/files/CANSO%20Articles%20of%20Association%20-%20Edition%2017%20-%202016%20.pdf)
- CANSO. (2020). *CANSO ATM Traffic Analysis Report* (Issue July).
- Cichoncki, P., Milar, T., Grance, T., & Scarfone, K. (2012). Computer Security Incident Handling Guide. In *NIST Special Publication* (Vols. 800–61). <https://doi.org/10.6028/NIST.SP.800-61r2>
- CIO. (2021). *The RACI matrix: Your blueprint for project success*. <https://www.cio.com/article/2395825/project-management-how-to-design-a-successful-raci-project-plan.html>
- EASA. (2017a). *European Centre for Cybersecurity in Aviation (ECCSA)*. EASA. <https://www.easa.europa.eu/eccsa>
- EASA. (2017b). Freddy Dezeure, Head of CERT-EU, explains the cooperation with EASA. *On Air*. <https://www.easa.europa.eu/newsroom-and-events/news/freddy-dezeure-head-cert-eu-explains-cooperation-easa>
- EASA. (2017c). Implementation of a European Centre for Cyber Security in Aviation(ECCSA). *On Air*, 14, 1–5. <https://www.easa.europa.eu/newsroom-and-events/news/implementation-european-centre-cyber-security-aviationeccsa#group-easa-related-content>
- EASA. (2020a). *Consolidated Annual Activity Report 2019 Of the European Union Aviation Safety Agency* (Issue June).
- EASA. (2020b). *What is CERT-EU, what is its role?* <https://www.easa.europa.eu/faq/24266>
- ENISA. (2017). *Improving recognition of ICT security standards Recommendations for the Member States for the conformance to NIS Directive* (Issue December). <https://doi.org/10.2824/176720>
- ETSI. (2020). *TR 103 306 - V1.4.1 - CYBER; Global Cyber Security Ecosystem* (Vol. 1).
- Eurocontrol. (2012). *Manual for National ATM Security Oversight Manual for National ATM Security Oversight* (Issue October).
- Eurocontrol. (2020). *Eurocontrol*. <https://www.eurocontrol.int/about-us#member-states>
- EUROCONTROL. (2020a). *EATM-CERT*. Eurocontrol. <https://www.eurocontrol.int/service/european-air-traffic-management-computer-emergency-response-team>
- EUROCONTROL. (2020b). *RFC2350 : Expectations for Computer Security Incident Response*.
- FAA. (2018). *NextGen - SESAR State of Harmonisation* (Third). Publications Office of the European Union, 2018. <https://doi.org/10.2829/90536>
- FAA. (2019). *Air Traffic by the numbers*.
- FIRST. (2019). *June 2019 Computer Security Incident Response Team ( CSIRT ) Services Framework Version 2. 0* (Vol. 0, Issue June).
- FIRST. (2020). *Forum of Incident Response and Security Teams*. FIRST. <https://www.first.org/>
- Ford, J., Faghri, A., Yuan, D., & Gayen, S. (2020). An Economic Study of the US Post-9/11 Aviation Security. *Open Journal of Business and Management*, 08(05), 1923–1945. <https://doi.org/10.4236/ojbm.2020.85118>
- Francy, F. (2015). *The Aviation Information Sharing and Analysis Center ( A-ISAC )* (Issue April).
- Goodwin, C., Nicholas, J. P., & Mckay, A. (2015). A framework for cybersecurity information sharing and risk reduction. In *Microsoft*.
- Hámornik, B. P., & Krasznay, C. (2017). Prerequisites of Virtual Teamwork in Security Operations Centers: Knowledge, Skills, Abilities and Other Characteristics. *Uni-Nke.Hu*, 16(3), 73–92. [https://folyoirat.ludovika.hu/index.php/aarms/article/view/1553%0Ahttps://www.uni-nke.hu/document/uni-nke-hu/AARMS\\_2017\\_03\\_05Hamornik\\_Krasznay.pdf](https://folyoirat.ludovika.hu/index.php/aarms/article/view/1553%0Ahttps://www.uni-nke.hu/document/uni-nke-hu/AARMS_2017_03_05Hamornik_Krasznay.pdf)
- HS SEDI. (2014). *Threat Intelligence using STIX and TAXII*. <https://secure360.org/wp-content/uploads/2014/05/Threat-Intelligence-Sharing-using-STIX-and-TAXII.pdf>
- International, C. (2020). *Why Cyber Criminals are Targeting the Aviation Industry*. Cyber Risk International. <https://cyberriskinternational.com/2020/04/06/cyber-threats-to-the-aviation-industry/>
- Killcrece, G. (2005). Incident Management. In *Software Engineering Institute*.
- Lykou, G., Iakovakis, G., & Gritzalis, D. (2019). Aviation cybersecurity and cyber-resilience: Assessing risk in air traffic management: Theories, Methods, Tools and Technologies. *Advanced Sciences and Technologies for Security Applications*, January, 245–260. [https://doi.org/10.1007/978-3-030-00024-0\\_13](https://doi.org/10.1007/978-3-030-00024-0_13)

**Faith Lekota and Marijke Coetzee**

- Majka, A., & Wacnik, P. (2019). Cooperation in aviation beyond Europe's Borders. *8TH EUROPEAN CONFERENCE FOR AERONAUTICS AND SPACE SCIENCES (EUCASS)*, 1–12. <https://doi.org/10.13009/EUCASS2019-852>
- Mana, P. (2019). EUROCONTROL's view on cyber risk, threats and challenges in ATM. *Cybersecurity and Resilience Symposium*, 25.
- Mana, Patrick, & Friligkos, V. (2019). EUROCONTROL / EATM-CERT SERVICES - SUPPORTING AVIATION TO BETTER MANAGE CYBER THREATS. *Integrated Communications, Navigation and Surveillance Conference (ICNS)*, 1–15. <https://doi.org/10.1109/ICNSURV.2019.8735282>
- Maybury, M. (2019). Information Sharing to Secure Cyberspace. In A. Armando, M. Henauer, & A. Rigon (Eds.), *Next Generation CERTs* (p. 120). IOS Press B.V. <https://doi.org/10.323/NICSP54>
- MITRE. (2020). MITRE. <https://www.mitre.org/>
- NIST. (2020). *National Institute of Standards and Technology*. NIST. <https://www.nist.gov/>
- Nolan, A. (2015). Cybersecurity and information sharing: Legal challenges and solutions. In *Cybersecurity and Cyber-Information Sharing: Legal and Economic Analyses*.
- OASIS. (2020). *OASIS-Open standards and Open Source*. OASIS. <https://www.oasis-open.org>
- Rechner, R., Whitlock, S., & Francy, F. (2012). Securing Airline Information on the Ground and in the Air. *Boeing*, 24–28. [WWW.Boeing.com/Boeingedge/aeromagazine](http://WWW.Boeing.com/Boeingedge/aeromagazine)
- SITA. (2021). SITA. <https://www.sita.aero/about-us/>
- Skierka, I., Morgus, R., Hohmann, M., & Maurer, T. (2015). CSIRT Basics for Policy-Makers. *Researchgate*, May 2015, 29.
- Wang, X., Miao, S., & Tang, J. (2020). Vulnerability and resilience analysis of the air traffic control sector network in China. *Sustainability (Switzerland)*, 12(9), 1–18. <https://doi.org/10.3390/su12093749>

# Biocyberwarfare and Crime: A Juncture of Rethought

Xavier-Lewis Palmer<sup>1</sup>, Ernestine Powell<sup>2</sup> and Lucas Potter<sup>1</sup>

<sup>1</sup>Biomedical Engineering Institute, Department of Engineering and Technology, Old Dominion University, Norfolk, USA

<sup>2</sup>Department of Neuroscience, Christopher Newport University, Newport News, USA

[xpalm001@odu.edu](mailto:xpalm001@odu.edu)

[ernestine.powell.12@cnu.edu](mailto:ernestine.powell.12@cnu.edu)

[lpott005@odu.edu](mailto:lpott005@odu.edu)

DOI: 10.34190/EWS.21.073

**Abstract:** The existence of BCS (Biocybersecurity), alternatively known as Cyberbiosecurity (CBS), as a hybrid field has been established over the past few years that explores vulnerabilities created at cyber-bio and cyber-physical intersections. Institutional leads, like Murch and DiEuliis (2019), have set about mapping the enterprise, uncovering a wide variety of vulnerabilities affecting the numerous cyber-physical and bio-digital vulnerabilities in the fields that comprise it (Berger, 2020; DiEuliis, 2020). Scholars like George (2019) and Wang (2020), have discussed the national security implications of the field, in addition to groups such as the Blue Ribbon, an American Bipartisan Commission on Biodefense, who have started assessing the risks where biology and cyber technologies converge as of last year in terms of national biodefense (Evans and Selgelid, 2015; George, 2019; Jefferson et al, 2014; Riley, 2019; Wang, 2020). This is not exhaustive and it is accurate to say that there is much to define and address within the wide map that has been drawn. One of which is the proper delineation between biocyberwar and biocybercrime. As it is, this small but growing field exists at the nexus of cybersecurity and biological sciences, where cyber, cyber-physical, and biosecurity meet (Murch et al, 2018). These, by large, are owed to the manifold improvements or creations of improvements in biotechnology, biomedical engineering, and adjacent technologies that BCS can affect. This paper aims to explore a possible delineation between biocyberwarfare and biocybercrime in order to and start the conversation before the technologies sufficiently catch up.

**Keywords:** biocybersecurity, cyberbiosecurity, cybersecurity, crime, biosecurity

---

## 1. Introduction

Imagine a dark alley, a pair of people, and a single weapon. The transpiring of violence between the parties and objects can be seen as one of a crime. Scaling the numbers of assailants, in a field of battle with thousands of people with guns, tanks, planes, and more at their disposal, few would question that this is the scene and context of a war. However, as our modern world develops, the acts that constitute an act of war and a mere crime have become more and more nebulous. The mere finesse of policy, lawyering, technological improvements, and more have changed much about how war is declared and conducted. Long gone are the days when wars were decided by geographic frontiers or had formal declarations or proclamations given before they began. And even now, the forms that warfare is hypothesized to take are questionable. Any work that dared to ask the question “What act or acts in our modern world should or do constitute an act of war, and which a mere crime?” would be ungainly and at least a little bit frustrated. This frustration could be extended to cyberwarfare and cybercrime. Instead, let the focus be on a small yet growing arena of crime – biocybersecurity (BCS).

The existence of BCS, alternatively known as cyberbiosecurity (CBS), as a hybrid field has been established over the past few years. Institutional leads, like Murch and DiEuliis (2019), have set about mapping the enterprise, uncovering a wide variety of vulnerabilities affecting the numerous cyber-physical and bio-digital vulnerabilities in the fields that comprise it (Berger, 2020; DiEuliis, 2020). Groups like Blue Ribbon, an American Bipartisan Commission on Biodefense, along with scholars like George (2019), Wang (2020), and more, among a growing field have discussed the national security implications of the risks where biology and cyber technologies converge (Evans and Selgelid, 2015; George, 2019; Jefferson et al, 2014; Millett, 2019; Riley, 2019; Wang, 2020). It is accurate to say that there is much to define within the wide map that has been drawn. One additional area that can use additional discussion is the line between biocyberwar and biocybercrime or where they converge. This small but growing field exists at the nexus of cybersecurity and biological sciences, where cyber, cyber-physical, and biosecurity meet (Murch et al, 2018). These, by large, are owed to the manifold improvements or creations of Improvements in biotechnology, biomedical engineering, and adjacent technologies that BCS can affect such as:

- Cheap and Efficient Genetic Sequencing (Bio to Digital Conversion)
- Digitalization of Medical Records and creation of open-source digital ancestry registries

- Creation of digitalized biotechnological resource records
- Biometric Authentication (Fingerprint, Face, Retina, Gait, etc.)
- Bio-Computing (Logic Circuits with Cells/Cell Components)
- Improved and Improving Brain Computer Interfaces
- Sophisticated Medical Implants and Carry-ons
- Democracy of educational resources online
- Cheap and powerful Smartphones
- Lab Automation via machine learning
- Enclosed robotic spaces
- Improvements in ergonomics
- The emergence of soft-robotics and emergence of micro and nano robots
- Telehealth (aided by OTG add-ons such as otoscopes and more)
- Gene Modification (Tools like CRISPR/CAS9 – cheaper and more accessible)
- Data Storage in Within DNA and other Genetic Material)
- And more.

For those that appreciate learning by example, the following would all be considered, in the arena of BCS, threat, or criminal endeavors there remain a complex variety of threats as shown below:

- 1. Ransomware attacks on a hospital
  - a. IoT enabled insulin pumps (Banerjee, 2017; Harris, 2019)
  - b. holding hospital records captive until money is paid to release it (Riggi, 2021).
  - c. fraudulent sales of personal protective equipment (Bond et al., 2020)
- 2. Phishing or data extraction on a biological research center
- 3. Using machine learning to optimize biological agents for a future attack
- 4. Remotely crafted, upgraded, and delivered viruses
- 5. In-person Human Healthcare DDOS via false wearable reports
  - a. DDOS on bio-based sensing systems (Bernal et al, 2020)
- 6. Smuggling of State Secrets in Biological Media
- 7. Improved Ransomware Aided by Biomimicry
- 8. Epi/Genetic Logic Bombs
- 9. Increased Unauthorized Tracking/Barcoding (Ibrahim, 2020)
- 10. IOT Implant/Wearable Spying or Manipulation (Banerjee, et al, 2017)
- 11. Undermining of Biometric Authentication Improved
- 12. Less predictable indirect surveillance (Plants/microbes) and more.

For a timelier example, the vaccine production line for COVID-19 has multiple places that is ripe for a BCS threat. The original design data for the vaccine has already been leaked (Porter, 2021). The potential for the production of the vaccine to be violated or swamped with counterfeits is possible.

Some of these threats already exist theoretically in the aforementioned forms or different versions. Others are possible in the near future, while others are much farther out. Nonetheless, it is important to think about these threats and the forms in which they may emerge prior to, especially as artificial intelligence (AI) may amplify threat creation with the more data generated and made bare (Jordan et al, 2020). As a special note, questions of digital afterlives or the merging of biological and digital selves are a fascinating topic and worthy of note for concern in BCS threats but will be out of the scope of this paper although threats here can be imagined (Pauwels

and Denton, 2018). Towards discussing the uncanny valley in which society may find itself when pondering of where crime stops and where war begins, without a limited setting, four things need to be discussed:

- First: the scalability and scaling of BCS threats. In the original example, a mugging perpetrated by a single person was hinted at.
- *This, itself, is not a particularly scalable act, since to have enough muggings to constitute a crime, you would need more and more people. For each individual committing a mugging, the cost of the perpetration increases to untenable economic levels.*
- Second: The different magnitudes of threats, as well as their polymorphisms and transcendence.
- *Here the threats that might evolve as research in this field develops are discussed here.*
- Third: There are many different angles of attack in the realm of BCS.
- *In order to employ them as an effective means of facilitating war, there are different combinations of threats that could improve the scale, the targeting, or the overall effectiveness of an act in such a way as to make a BCS crime into an act of war.*
- Fourth: The legacy of potential BCS threats with respect to prior conflicts.
- *It is important to think of what these threats may mean with the passage of time and how a temporal focus may later implementation of said threats.*

An important note to be had is that discussion of one point will bleed into the others for purposes of discussion and demonstration of fluidity of discussion.

## **2. Section 1: Scalability**

An enormous amount of readers of this paper have probably had their data stolen in one way or another. A smaller subset have probably been active victims of identity theft. This act is a nuisance and is undeniably criminal, but at what point could it be considered an act of war? That is the primary question, and it can be addressed through the concept of scale. In the world of BCS, there are abilities to scale up what could easily be called nuisances into acts that cast doubt on long-held systems. Some crucial examples are ID theft for bank loans to the point that an entire nation's population credit is ruined, ransomware attacks on a hospital system (Special note: this magnitude of this attack could be particularly amplified if it targets military hospitals), conventional biological attacks (like the strategic spreading of viruses or bacteria on commonly used public services or in industrial lots or more, using drones and informatics amplified approaches to find the best spots for attack and spread. Within all attacks like the above, the targets can expand to a large number of individuals.

The last one deserves special mention. In the case of Covid-19, it has claimed more than 1 million lives, globally, while infecting tens of millions. While it is not the most devastating natural pathogen, in regard to the casualties of other historic threats like the 1918 Flu or the Black Plague, it has shown that modern countries still have trouble containing such threats, for a variety of reasons (Mueller, 2020; Nakamoto et al, 2020). The lesson that this is normally just a biosafety threat demonstrates is that malicious actors can make up for a lack of nukes, through exploiting country-wide ignorance of biology, epidemics, and other aspects of health safety, combined with the increasing accessibility of biotechnological hardware, thus cyber-physical interfaces, that allow for customized and distributed biotreats, if so desired. In the time since nations have banded together to tackle Covid-19, multiple labs have been able to isolate and create weaker, alternative versions to allow the study of Covid-19. A mix of academic and DIY Biology groups like JOGL, RADVAC, and other teams have shown that vaccines and testing means for Covid-19 can be constructed with intermediate effort and knowledge (Caplan & Bateman-House; Moritz, 2020; Tennant et al, 2020). The existence of tools like CRISPR and CAS9, as well as cheaper variants for less than \$200, and dropping, along with an increasing array of open source software and robotic hardware have lowered the barriers of creating customized threats than can be delivered through cheap drones (la Cour-Harbo, 2017; May, 2019; Pylatiuk et al, 2018; Zegart, 2020). Much like with how Iran and North Korea have abandoned conventional means of matching the US and the rest of the West in favor of asymmetric fighting potential, such as through cyber warfare, it is not unlikely that they have or already are considering how to exploit this potential in BCS to add their arsenal.

When a country attacks using its cyber arsenal, it is by large for respect or financial incentives, and thus on a criminal basis. However, the adoption of BCS could change that, in which each act can get swiftly elevated to a wartime act, and that could be preferred as BCS threat capacity could allow it to weaken another while claiming

innocence. There is a growing question of how the nature of attribution changes in these attacks as with cyber. One could track DNA and other signatures, but there is a proper question with regards to a new game of cat and mouse with users creating new facilities and means for designing and threat and removing the trace of their involvement, as well as what does this mean in the broader sense for society (Hester, 2020; Weber and Kämpf, 2020). This is almost beside the point of the original pillar but is worth discussion as it can change perceptions of scales affected; in practical terms this could mean one party only identifying or properly attributing a fraction of acts against it, in addition to new means of framing and social engineering for false attribution. The future of BCS threats in scales of attack is one that is still maturing and difficult to predict.

### **3. Section 2: Threat magnitudes, polymorphisms, and transcendence**

Related are the threat magnitudes. This concerns not just the number, but the scales and platforms through which government, and thus military power is disrupted. For example, in terms of the magnitude of effect that would compromise a town or city affecting 1000 people, it can yield different results depending on the specificities and locality. For a small, rural town or research village, this can wipe out local power and capacity. However, this same number is a drop in the bucket towards affecting a city with a colossal population that is spread out, like that of Los Angeles, California, or the greater Chicago metropolitan area. That same 1,000 might have a markedly higher effect on local power in a highly dense area such as San Francisco, California, or New York City, New York. Nations trying to defend geographically incongruent territories, with incongruent populations to match might face heightened difficulty. This not only raises the importance of smart cities that can provide enhanced metrics on areas but also to why urban and rural planning does not have a heightened space in defense budgets versus mere munitions; the effect of attacks of smart agriculture or vertical farms in cities is obvious (van der Linden et al, 2020).

In terms of polymorphisms, this concerns not just the angles from which a threat can arise, but the many forms in which one threat can have. For example, a BCS threat in the form of a compromised IoT arm prosthetic presents multiple threats, assuming that a vulnerability in reading movement commands for the arm, as well as controlling it, is found. On the scale of just an individual, what this means is that a malicious actor might use this information to spy on the actions of the user throughout parts of the day, take control of the arm at various times, and or use indirect data to spy on others through mapping communications in typing or writing motions, and or spying on other users for whom this hand comes in contact with. For instance, more advanced versions of commonly used devices, such as insulin pumps and Holter monitors, have the ability to track location, save biometric data, and even have access to WiFi or the ability to be used with a mobile phone (Banerjee, 2017). These allow for patients to have individualized access to their data and better tracking for their knowledge, as well as their healthcare team, but it also allows for unethical hackers to tap into that data as well and use it however they please (Banerjee, 2017). Consider if these devices also have the capabilities to handle voice commands, as in the case with some insulin pumps, or are involved in movement, as in the case of the formerly mentioned arm prosthetic, there can be heightened risks involved. For example, when an individual with those devices is in a medical or legal capacity, being in proximity to health or legal information can be met with wide-ranging consequences, especially if the individual is involved with the discussion of high stakes information. This is already taken into account with many high security clearance level jobs in the military that ban the use of cell phones and certain versions of Holter monitors while actively performing one's role on base. On this individual level alone, multiple heinous crimes can arise, but this becomes more serious when the vulnerability that allows for this means of spying and manipulation is used en-masse by a state-level actor. Concerning current cases, the U.S., Russian, and Chinese government are some of the few that have regulated which cell phones are banned for use by those serving in their military (Nichols, 2018; BBC, 2019; Rempfer, 2020). The U.S. military in particular tends to have more regulations when deploying in certain regions of the world, such as some army units to the Middle East ( Nichols, 2018; BBC, 2019; Rempfer, 2020). Aside from these examples, further mass violations of a nation's sovereignty can occur, leading to vast volumes of data gained that can result in meaningful changes to power balances between nations, as well as the possibility of physical damages to the state depending on the extent of spying and manipulation. Through the lens of popular countries of study, one use could mean the spying of communications through that for influencing political outcomes. Another use could be for gaining intellectual property about the arm, use metrics, and of what interfaces with said arm or the arm is given access to. Rogue state use could be for building an understanding for disrupting use by any opposing armed forces personnel who may have prosthetics, for an asymmetric counter. Even counted as a crime, such as access, and exploitation meaningfully could provoke a response as affected parties would likely see the act as one of war.



Lastly, the prior two sub-points bring us to the question of transcendence. BCS threats, by their hybrid nature, transcend the pointed nature of a standard biosecurity or cybersecurity threat because they often combine, amplify, or create new vulnerabilities that produce their own class ability as well as problems. A DNA-based attack on a forensics or medical device can give stealth or versatility to an attack that would otherwise be intercepted quickly if done through a purely electronic format such as phishing, DDoS, or deliverance of a viral payload (Riley, 2019; Ney, 2017). On the other side, in targeting an industrial biotechnology lab, a saboteur might find themselves modifying software that governs the processing of biologics, to allow inferior reagents to be used. Cases like these more or less require high-level knowledge, but the existence of a biocybertechnological equivalent of Metasploit may not remain an impractical fantasy, raising the question of how soon society may see sophisticated script kiddies in this domain emerge. In the meantime, vulnerability exploits appear to breach the threshold of deliberate instead of just for fun, given the specificity needed to pull them off. This leads to the possible opinion that actors at any nation state-level committing a crime at this capacity should be seen as committing an act of warfare given that the stakes of exploitation are high enough.

#### **4. Section 3: Confluences of threats**

Threats can be magnified through the proper targeting, but this is amplified when threats are combined. For example, a DNA attack, as one shown by Ney in 2017, on a forensics unit equipped with a sequencer can serve as a multi-pronged tool of recon and compromise (Ney et al, 2017). Attacks on ancestry databases can be used to not only take or fake data but cause secondary and tertiary effects on the victims through problems to encounter by third parties who service the affected parties with the data that is affected, such as professionals in insurance or history (Ney et al, 2020). Depending on the target, this might be seen as an act of war if the person of interest or one of the victims is or is linked to someone of national importance and if this attack is a state-based in origin. Similarly, someone flooding the National Center for Biotechnology and Industry, NCBI, could wreck not only the entries of biological information for an organism or biologic with national, medical importance, but it could serve as a means to attack an entire country, through agricultural and or pharmaceutical sabotage (Schabacker et al, 2019; Walsh and Steilein et al, 2020; Warmbrod et al, 2020). Famines of the 20th century, linked to agricultural mismanagement could in time be dwarfed. This would not necessarily be through mass mismanaged monoculture, but it could be through mass individualized and surgical sabotage of agricultural protocol due to record sabotage or calculated biofouling. This may in turn influence more countries to create their own versions of the NCBI and back reforms in internet sovereignty (Lindsay, 2015).

To avoid digression, BCS threats can be more numerous in creativity, and in time, they can be more destructive power through an increase in routes of attacks. One can, for example, use the NCBI as a platform for not only getting clues on how to attack a population but use it as a library for customizing their attacks beyond recognizable means of distribution at the origin. Further, attacks at this intersection can become stealthier and more surgical. Paired with these advantages, owing to mass ignorance in biology owing to under education and miseducation worldwide, these attacks improve in their degree of lucrative design and status. What this means for the average citizen or nation around the world to defend against these threats, especially those crafted by experience interdisciplinary teams, is quite poor. This is to say that the threat increases by not just a magnitude, but a dimension. With the increasing interconnectivity of citizens, one attack towards one person can mean an attack towards many more, intentionally, or unintentionally.

Crime needs not an intention nor is shielded under the guise of just following orders, in modern international law as shown by the Nuremberg Trials and more of last Century. However, injury, severe enough can prompt injured parties to engage in war against parties who fail to or unable contain criminal elements, under the banner of protecting national sovereignty. This may mean that BCS threats can cut away the practical need to differentiate between a crime and wartime act in some circumstances, as it can lead to war. One can think back to numerous special operations conducted by strong regional powers on weaker states to apprehend, confront, or neutralized suspected terrorist elements in less stable, secure countries, without their explicit consent. Groups engaging in BCS threats would be wise to consider how a criminal act of theirs could affect their nation's sovereignty.

#### **5. Section 4: The legacy of the threats**

It is time to consider the legacy of the threat. Let us take a look at the September 11 attacks on the Twin Towers and their later precipitation the Iraq War, and its later expansion into Afghanistan and others through the War on Terror. Public sentiment fueled by these attacks propelled long-term support for expanded military

engagements in the Middle East, Africa, and Asia, which cost the United States and allies trillions of dollars and thousands of lives, over a 20 year and growing period (Crawford, 2019). What this potentially translates to is that an enemy who can afford to suffer large losses can etch out a victory by asymmetric means in getting one's enemy to exhaust themselves in other means that are more damaging. Resources spent to combat one's enemy in the west have taken from resources to otherwise enhance and maintain it, like education, healthcare, and less pursued basic research that has historically led to breakthroughs in other areas. However, this view is more easily seen when an assumed uniformity in objectives between sides is not taken for granted. In terms of resources and values, envisioned as a game of chess, the proverbial chessboard is not a mirror in the middle.

Each side does not have the same pieces, plays by the same rule set, nor assigns the same value to or cares about the same pieces. Further, each piece is uniformly one-piece or translatable to that of the other side. The terrorists and or defending forces are not equivalent to that of the liberators and or the attackers, in values or resources. Why should one society think that successes against different foes will carry the same meaning or legacy? Much like the US failures in Vietnam, many defending forces were willing to sacrifice vast sums of life for its victory and the eventual US withdrawal, choosing to attack asymmetrically, when it was underpowered conventionally. This bears similarity to what is unfolding in Afghanistan wherein the Taliban are seeing US forces withdraw, after withstanding 20 years of assaults, lost territory, and lives while seeing magnitudes greater losses in resources on their enemies (Tariq et al, 2020). Similarities can be found in the on-going defense that North Korea has played on the West through ongoing peacetime provocation, nuclear capacity building, and maintenance of non-conventional forces, both physical and cyber, to hold out. These all concern matters of national sovereignty, and thus war, giving way to a legacy that will not escape history books or fail to leave legacies in policy and public thought.

In terms of BCS, attacks like that SARS in the early 2000s reported weaponization of Covid-19 by extremist groups, the potential bio-electronic industrial espionage, and more pose questions of how wide a net must cast in terms of adequate BCS defense in defending forward initiatives (Ackerman and Peterson, 2020; Cooper, 2006; Evans and Selgelid, 2015; George, 2019; Millet et al, 2019; Mueller, 2020). Further, it poses questions of how empowered smaller groups are with the improved accessibility of weaponized biocyber means. The legacy that is before us could be that small clusters of societies are viewed with more suspicion, that biological reagents and certain biotechnological hardware is more closely regulated, and further, that western societies welcome even more intrusion into the lives of individuals, as shown through successive Patriot Act and others, which notably heightened surveillance efforts (Hernandez, 2020). Provided that you, the reader, accessed this paper online, without adequate means of layered, obscuring methods, you are likely being monitored as well. These are things that societies must continue to grapple with as BCS weaponization moves from irregular and incongruent petty crime to regular and implemented means of warfare, side by side with cyberwarfare. Both crimes and war evolves and their scale and magnitude pale to the horrors that warfare elevates their base act to. One must be careful in our delineations, means of deterrence, and means of containment.

## **6. Conclusion**

This paper is one of many growing topics from within BCS and serves to promote discussion of the delineation, separation, and meaning of BCS in terms of criminality and war capacity. In it, varieties of potential BCS attacks were discussed, along with the how of advances that enabled them. Further, a discussion was given towards the scalability and scaling of BCS threats, different magnitudes of threats, the many different angles of attack in the realm of BCS, polymorphism, and transcendence of threats, and the legacy of BCS threats concerning prior conflicts. Altogether, these promote the question of how deeply society is thinking of how BCS threats affect the perception of malicious actions. The question of what makes the difference is obvious, as scenarios are context-dependent, but the transient question that emerges, that is, "Of what elements should matter most?", is beyond the scope of this paper. As BCS matures, intake of the field must be open, thoughtful, and comprehensive, for societies aiming to grapple with the future that they will create.

## **References**

- Ackerman, G., & Peterson, H. (2020). Terrorism and COVID-19. *Perspectives on Terrorism*, 14(3), 59-73.
- BBC News (2019, February 20). Russia bans smartphones for soldiers over social media fears. Available at <https://www.bbc.com/news/world-europe-47302938>.
- Banerjee, S. S., Hemphill, T., & Longstreet, P. (2017). Is IOT a threat to consumer consent? The perils of wearable devices' health data exposure. *The Perils of Wearable Devices' Health Data Exposure (September 18, 2017)*.

- Bond, S., Romo, V., and Wamsley, L. (2020, October 29) U.S. Hospitals Targeted in Rising Wave of Ransomware Attacks, Federal Agencies Say. Available at <https://www.npr.org/2020/10/29/928979988/u-s-hospitals-targeted-in-rising-wave-of-ransomware-attacks-federal-agencies-say>
- Berger, K. M. (2020). Addressing Cyber Threats in Biology. *IEEE Security & Privacy*, 18(3), 58-61.
- Bernal, S. L., Martins, D. P., & Celdrán, A. H. (2020, June). Distributed Denial of Service Cyberbioattack Affecting Bacteria-based Biosensing Systems. In *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 279-282). IEEE.
- Caplan, A. L., & Bateman-House, A. (2020). The danger of DIY vaccines.
- Cooper, M. (2006). Japanese tourism and the SARS epidemic of 2003. *Journal of Travel & Tourism Marketing*, 19(2-3), 117-131.
- Crawford, N. C. (2019). United States budgetary costs and obligations of post-9/11 wars through FY2020: \$6.4 trillion. *Watson Institute for International and Public Affairs, Brown University*.
- DiEuliis, D. (2020). Parsing the Digital Biosecurity Landscape. *Georgetown Journal of International Affairs* 21, 166-172. doi:10.1353/gia.2020.0031.
- Evans, N. G., & Selgelid, M. J. (2015). Biosecurity and open-source biology: the promise and peril of distributed synthetic biological technologies. *Science and engineering ethics*, 21(4), 1065-1083.
- George, A. M. (2019). The national security implications of cyberbiosecurity. *Frontiers in Bioengineering and Biotechnology*, 7, 51.
- Harris, B. (2019, July 01). FDA issues new alert ON Medtronic insulin pump security. Retrieved from <https://www.healthcareitnews.com/news/fda-issues-new-alert-medtronic-insulin-pump-security>
- Hernandez, S. (2020). Surveillance Technology Toward a Dystopian Future.
- Hester, R. J. (2020). Bioveillance: A Techno-security Infrastructure to Preempt the Dangers of Informationalised Biology. *Science as Culture*, 29(1), 153-176.
- la Cour-Harbo, A. (2017). Mass threshold for 'harmless' drones. *International Journal of Micro Air Vehicles*, 9(2), 77-92.
- Ibrahim, M., Liang, T. C., Scott, K., Chakrabarty, K., & Karri, R. (2020). Molecular Barcoding as a Defense against Benchtop Biochemical Attacks on DNA Fingerprinting and Information Forensics. *IEEE Transactions on Information Forensics and Security*.
- Jefferson, C., Lentzos, F., & Marris, C. (2014). Synthetic biology and biosecurity: challenging the "myths". *Frontiers in public health*, 2, 115.
- Jordan, S. B., Fenn, S. L., & Shannon, B. B. (2020). Transparency as Threat at the Intersection of Artificial Intelligence and Cyberbiosecurity. *Computer*, 53(10), 59-68.
- Lindsay, J. R. (2015). Exaggerating the Chinese Cyber Threat. *Quarterly Journal: International Security*. Retrieved from <https://www.belfercenter.org/sites/default/files/leaqcy/files/lindsay-china-cyber-pb-final.pdf>.
- May, M. (2019). A DIY approach to automating your lab. *Nature*, 569(7754), 587-589.
- MILLETT, K. K., dos Santos, E., & MILLETT, P. D. (2019). Cyber-Biosecurity Risk Perceptions in the Biotech Sector. *Frontiers in bioengineering and biotechnology*, 7, 136.
- Moritz, R. L., Berger, K. M., Owen, B. R., & Gillum, D. R. (2020). Promoting biosecurity by professionalizing biosecurity. *Science*, 367(6480), 856-858.
- Mueller, S. (2020). Facing the 2020 Pandemic: What does Cyberbiosecurity want us to know to safeguard the future?. *Biosafety and Health*.
- Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., & Peccoud, J. (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bioeconomy. *Frontiers in bioengineering and biotechnology*, 6, 39.
- Nakamoto, I., Wang, S., Guo, Y., & Zhuang, W. (2020). A QR Code-Based Contact Tracing Framework for Sustainable Containment of COVID-19: Evaluation of an Approach to Assist the Return to Normal Activity. *JMIR mHealth and uHealth*, 8(9), e22321.
- Ney, P., Koscher, K., Organick, L., Ceze, L., & Kohno, T. (2017). Computer Security, Privacy, and {DNA} Sequencing: Compromising Computers with Synthesized {DNA}, Privacy Leaks, and More. In *26th {USENIX} Security Symposium ({USENIX} Security 17)* (pp. 765-779).
- Ney, P., Ceze, L., & Kohno, T. (2020). Genotype extraction and false relative attacks: security risks to third-party genetic genealogy services beyond identity inference. In *Network and Distributed System Security Symposium (NDSS)*. New York: NDSS.
- Nichols, H. (2018, May 2). Pentagon says Chinese cellphones are 'security risk,' bans sale at bases. Available at <https://www.nbcnews.com/news/military/pentagon-says-chinese-cellphones-are-security-risk-bans-sale-bases-n870756>.
- Pauwels, E., & Denton, S. W. (2018). The internet of bodies: life and death in the age of AI. *Cal. WL Rev.*, 55, 221.
- Porter, S. (2021, January 14). Pfizer COVID-19 Vaccine data leaked by hackers. Retrieved from <https://www.healthcareitnews.com/news/emea/pfizer-covid-19-vaccine-data-leaked-hackers>
- Pylatiuk, C., Vogt, M., Scheikl, P., & Gottwald, E. (2018, July). Automated Versatile DIY Microscope Platform. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5310-5312). IEEE.
- Rempfer, K. (2020, January 6). No cellphones, laptops were allowed to go with Army 82nd paratroopers deploying to Middle East. Available at <https://www.armytimes.com/news/your-army/2020/01/06/no-cell-phones-laptops-were-allowed-to-go-with-82nd-paratroopers-deploying-to-middle-east/>

- Riggs, J. (2021). Ransomware Attacks on Hospitals Have Changed. Available at <https://www.aha.org/center/cybersecurity-and-risk-advisory-services/ransomware-attacks-hospitals-have-changed>.
- Riley, Kim. 2019. "Bipartisan Commission on Biodefense Considers Cyberbiosecurity Threats." *Homeland Preparedness News*, September 23. Available at <https://homelandprepnews.com/countermeasures/36927-bipartisan-commission-on-biodefense-considers-cyberbiosecurity-threats>.
- Schabacker, D. S., Levy, L. A., Evans, N. J., Fowler, J. M., & Dickey, E. A. (2019). Assessing cyberbiosecurity vulnerabilities and infrastructure resilience. *Frontiers in bioengineering and biotechnology*, 7, 61.
- Tariq, Muhammad, Muhammad Rizwan, and Manzoor Ahmad. "US Withdrawal from Afghanistan: Latest Development and Security Situation (2020)." *sjesr* 3.2 (2020): 290-297.
- Tennant, J., Francuzik, W., Dunleavy, D. J., Fecher, B., Gonzalez-Marquez, M., & Steiner, T. (2020). Open Scholarship as a mechanism for the United Nations Sustainable Development Goals.
- van der Linden, D., Michalec, O. A., & Zamansky, A. (2020). Cybersecurity for smart farming: socio-cultural context matters. *IEEE Technology and Society Magazine*.
- Walsh, M., & Streilein, W. (2020). Security Measures for Safeguarding the Bioeconomy. *Health security*, 18(4), 313-317.
- Wang, X. (2020). COVID-19 Epidemic and Enhancing China's National Biosecurity System. *Journal of biosafety and biosecurity*.
- Warmbrod, K. L., Trotochaud, M., & Gronvall, G. K. (2020). Shaping the US Bioeconomy for Future Economic Development and Sustainability. *Health security*, 18(4), 265-266.
- Weber, J., & Kämpf, K. M. (2020). Technosecurity Cultures: Introduction.
- Zegart, A. (2020). Cheap fights, credible threats: The future of armed drones and coercion. *Journal of Strategic Studies*, 43(1), 6-46.

# Matters of Biocybersecurity With Consideration to Propaganda Outlets and Biological Agents

Xavier-Lewis Palmer<sup>1,3</sup>, Ernestine Powell<sup>2</sup> and Lucas Potter<sup>3</sup>

<sup>1</sup>School of Cybersecurity, Old Dominion University Norfolk, USA

<sup>2</sup>Department of Neuroscience, Christopher Newport University, Newport News, VA, USA

<sup>3</sup>Biomedical Engineering Institute, Department of Engineering and Technology, Old Dominion University, Norfolk, USA

[xpalm001@odu.edu](mailto:xpalm001@odu.edu)

[ernestine.powell.12@cnu.edu](mailto:ernestine.powell.12@cnu.edu)

[lpott005@odu.edu](mailto:lpott005@odu.edu)

DOI: 10.34190/EWS.21.085

**Abstract:** The modern era holds vast modalities in human data utilization. Within Biocybersecurity (BCS), categories of biological information, especially medical information transmitted online, can be viewed as pathways to destabilize organizations. Therefore, analysis of how the public, along with medical providers, process such data, and the methods by which false information, particularly propaganda, can be used to upset the flow of verified information to populations of medical professionals, is important for maintenance of public health. Herein, we discuss some interplay of BCS within the scope of propaganda and considerations for navigating the field.

**Keywords:** biocybersecurity, cyberbiosecurity, cybersecurity, public health, biosecurity

---

## 1. Introduction

The flow of information to a medical professional is variable in terms of content and volume. Generally, licensed healthcare providers and allied health professionals are supplied a myriad of information during their education and training. While the number of years and tests required for employment varies amongst the possible careers for this population, generally there are one or multiple certifications or licensure exams at the terminus of their formal education. Additionally, there tend to be required clinical internships, often before those formerly mentioned exams, that allow these professionals to officially join the workforce. This is especially the case for professions that require the internship as part of the requirements to legally join the field. Even after fully earning the title of their profession, the learning continues in the form of a specific number of continuing education credits, announcements from supervisory medical professionals or healthcare management, specialized public health announcements conversations with their peers, and entries in academic medical journals. In contrast, the general population, specifically patients without a formal medical background, instead utilize a much wider front of knowledge. This front encompasses newspapers, television programs, videos, and posts on social media from healthcare professionals and those without any medical training, and official announcements from government and non-government agencies. Thus, information sources are available to everyone, regardless of medical background. However, people with limited or no medical background are exposed to this often with scientifically unsupported information void of the knowledge helpful to separate facts from misinformation or disinformation. This problem uniquely rests at the intersection of biosecurity, cyber-physical security, and cybersecurity, wherein the interlock of biology in the form of data or as a pathological agent especially in times of pandemics, can be used to exploit systems. Governments and economic entities would find value in preventing exploits that could disrupt their stability. In this article, we seek to suggest how avenues of knowledge, both for the medical professional and the layperson, may be hindered or compromised regarding an indeterminate biological agent.

## 2. Veracity of information

We will begin by discussing how laypeople obtain information. As outlined above, their methods are far less formalized than that of medical professionals. For example, in the U.S., the ability for news services to operate under FCC regulations prevents intentionally broadcasting false information under the guise of news (Broadcasting False Information 2021). However, this is predicated on the broadcaster completely understanding the nature of biological agents and such information is variably available. Online services (either sites that are commonly referred to as "Social Media", but any network that allows for information sharing - which could even be a video game service provided it enables the communication of citizens) are given some protection from what their users choose to share under Section 230 of the Communications Decency Act

(Section 230 of the Communications Decency Act). These services may have internal policies that support the dissemination of truthful information. However, they are not required to do so given lack of consequences for transmission.

### **3. Exploitation of a general public**

Many methods exist for exploiting communities through generating or disseminating data. Some routes obfuscate facts, misdirect, cause dissemination of false information, and unnecessary emphasis in reporting to disturb public focus. These can all be directed and gamed by state-level and local actors for various motives, including but not limited to applying economic pressure to another state to weaken, probe, or use it as a testbed for in the form of a deferred study, or a combination. In the case of a deferred study: this may not be uncommon. For example, 90% of experimental drugs were reported to have been tested outside of the United States and Canada in 2017, and further, these drugs commonly yielded more favorable results in developing countries, which may hint at questionable motives in data collection and potential to mislead those not privy to eccentricities in studies undertaken (Panagiotou, et al, 2013; Robbins, 2017). This route can open the door to indirect probing of nation health, while also being used as a vehicle for misleading interested parties in said data. Albeit this potential being fringe, the consideration can help map future scenarios regarding the discussion of international testing dynamics in epidemiological interventions.

For simplification, an episode of the COVID-19 pandemic will be used as an example in considering inherent dangers. One potential example of this can be found from hypothetical imaginings based on praise and critique in the coverage of Sweden, who resisted lockdown measures in favor of their economy, but eventually changed its policy (Ellyatt, 2020). Data in Sweden has since been compared with data of other countries that both chose different lockdown strategies (Taylor, 2020). There is no evidence for one to say that Sweden's decision was affected by disinformation, but it is possible that select episodes of Sweden's policy could be misrepresented for propaganda to encourage harmful or counter-productive healthcare policy in other nations, but this remains to be seen. For consideration, the closest example might be through the encouragement of hydroxychloroquine use in Brazil as influenced by the US in 2020, but it must be stated that there's inconclusive evidence to imply that the use and encouragement were malicious, especially at a state-level, and this is just as a thought exercise (Berlivet and Löwy, 2020). Early on and following experiments for verification, some researchers expressed concern due to inconclusive evidence of hydroxychloroquine's efficacy and or the negative effect that stockpiling had on patients of other maladies who had a more immediate need of the drug in other countries, profits notwithstanding (Cavalcanti et al, 2020; Falcão et al, 2020; Martins-Filho et al, 2020; Palmeira et al, 2020). This illustrates a potential for propaganda in the form of disinformation or misinformation to deliver an effect in a country at cost but in favor to others. If eventually and truly acted upon, this means of BCS propaganda is important to monitor.

### **4. The routes in further detail**

Obfuscation involves propaganda that makes meaningful information difficult by flooding the public with useless or inaccessible information. This can come in the form of media that recklessly reports threat severity and the non-solutions, ultimately resulting in much of the public rejecting mainstream news advice delivery. Non-compliance remains a risk ((Feunekes and Hermans, 2020; Nivette et al, 2020). In terms of COVID-19, this has taken the form of news that worn the public on information concerning origins, validity, and impact of COVID-19, as well as any proposed intervention, such as stay-at-home orders or lockdown encouragement, social distancing, mask-wearing, and vaccine reception -- obfuscation typically complicates relief efforts (Forman et al, 2020; Scheid et al, 2020). Furthermore, such obfuscation can empower acceptance of ineffective and insufficiently evidenced interventions internationally, such as unnecessarily wide mesh masks that cannot stop droplet transmission, herbal concoctions, poorly-tested pharmaceuticals and other misinformed efforts of treatment, such as encouraged contraction. Such propaganda in essence carries hints of community and state capture as evidenced in the widespread and rapidly increasing cases of death and injury in countries where prominent leaders have resisted scientific consensus or peddled unproven treatments (Freckelton, 2020; Nordling, 2020; Reihani, 2020).

Another BCS-propaganda route is misdirection, which is related to obfuscation but differs on the directionality of consumer attention and effort and can lead to scapegoating. In the case of COVID-19, many individuals directed considerable animosity towards people of Asian descent in the early months of 2020 (Misra et al, 2020; Gover et al, 2020; Ziems et al, 2020). Other examples of misdirection can be found in the misplaced emphasis

on the logistics of solutions by competing political parties as evidenced through partisan bickering over aid and relief packages in legislation (Bard, 2017; Fiedler, 2020; Nicola et al, 2020). One more example can be found in conspiracy theories channeling dissatisfaction and action towards governments with accusations of liberties being stripped away and even protested efforts to mitigate the spread of the virus. (*Coronavirus lockdown protest: What's behind the US demonstrations?*, 2020). Malicious state-level actors looking to reduce pressure on their nations through inciting internal chaos can find such BCS-propaganda to their benefit. Some may use the information to prey on relief groups focus on less-privileged areas. Consequently, many health-focused groups are looking for more accessible means in the face of short supplies for the underserved who are likely to face delays with access to the vaccine in ways similar to the early days of the AIDS epidemic, which also amplified on-going disparities (Thrasher, 2020; *Pushing for a People's Vaccine for COVID-19*, 2020).

Concerning perhaps the most obvious propaganda, disinformation, involves the transfer or circulation of verifiably incorrect information. Disproving false information, amid widely imperfect public perception, can prove difficult. While spreading false information can confer many benefits for a malicious actor, it can be challenging to implement it without debating the proliferation of science communicators and general fact-checkers. However, especially highly-partisan, and conspiracy-aligned individuals can act as prominent vectors for spreading false information (Wallace-Wells, 2020). Additionally, research has demonstrated that false information spread by bots has low penetrance, but emotionally-charged, human-spread information can exceed transfer rates of truthful news (Langin, 2018). Thus, by using low-information, low-disciplined people as vectors for false information, a seed can be sown that persistently undercuts control efforts to defeat pandemics and related issues. However, it is worth restating that plenty of high-information, highly educated people can be successfully targeted as well. As others have alluded, the public, by and large, is highly susceptible, bringing complication to plans that tackle just education as a sole defense within info-war environments (Aro, 2016; Woolley and Howard, 2018).

A final example is that of omission, which can be accomplished via intentionally underreporting instead obscuring health-related data through direct attacks of networks that hold records. Clues can be found in disputes over Florida's data covering their number of Covid-19 cases, which some allege that information related to COVID-19 deaths or cases has been inaccurately reflected or hard to properly access at times (Chacin & Klas, 2020; *Why are coronavirus deaths doubling in Florida's nursing homes?*, 2020). While there may not be malfeasance, this controversy points to a potential for a malicious actor in a hypothetical situation to tailor propaganda that encourages state-level omissions. Questions already exist in other countries with regard to true COVID-19 tolls (Richards, 2020). This can create a false sense of business-as-usual and lead to the early lifting of travel bans and the potential of these locations becoming hotspots for areas near other parts of the country via travel.

## **5. Disinformation for medical professionals**

Medical professionals have an oath and strong legal and career consequences that help to prevent the intentional spread of false information. These consequences include revocation of one's license or certification to work in their profession, difficulty obtaining another license or certification in the medical field, or lawsuits with steep financial consequences. Even the publication of medical information requires some standards which could be called exacting, such as the accepted standard double-blind, peer-review process for academic journal integrity. However, this standard has been compromised to a degree by predatory publishing practices wherein quality and the information supporting a given work are not up to sufficient thresholds. Theoretically, the violation of academic medical publishing can also be considered a BCS-threat as it involves an exploit of cyber-systems that pertain to matters of biosecurity. Two of the authors, Palmer and Potter discussed this previously at their presentation, "Commentary on Cataloguing Biological Assets and Academic DDoSing: Thoughts on Biocybersecurity in the Global South" at a prior conference, with research on-going (Palmer and Potter, 2021). Hypothetical scenarios considered, one route of violation is through the DDOS of reviewers, in which malicious actors can inflict damage internationally through means of slowing meaningful review and publication of data on biological assets, which could lead to fractured biomedical logs for medical communities, by targeting different geographic areas of different fields. For example, one could target one paper for surgeons and another for internal medicine practitioners, with unnecessarily, differing information. In essence, this could further contribute to the already notable differences in opinion of care among mainstream health practitioners. Another is through the skewed representation of national assets, through false data, disrupting the progress of developing nations who look to expand their economies through biological resource discovery. Another is

through the slowed global advancement of biologically-based research through a combination of the above; this of course has indirect effects on lives saved, as disruptions to a country's economic stability can impact the quality-of-life. Given that biodiversity is largely underestimated and lack of identification for existing pathogens, this could hamper pandemic-related research (Larson, Ghosal, & De Sousa, 2020; Pennisi, 2020). Already various scientists and other specialists are acting proactively to root out and pre-empt pandemic capable pathogens and such may see their efforts diluted if abuses of medical literature records are unchecked (Cox, 2021; Scientists focus on bats for clues to prevent next pandemic, 2020). Such requires biological database manipulation which connects to the interlock of biology, making this inescapably BCS.

BCS propaganda could also vary professional opinions and medical decision-making. This overall can affect patient healthcare plans, especially where healthcare teams require a combination of professionals. This route may still threaten given: a higher degree of training undergone by healthcare providers in charge of creating and finalizing patient care plans, increased curation of high-caliber research in reputable journals, and decreased use of less-reputable journals by practitioners. Yet, the barrier to sabotage prevention may be lowered through flooding of ranks in healthcare with skeleton crews due to staffing issues and consequent inadequacies in facility-specific training; decreased cognitive functioning in exhausted units; infiltration of ranks with ideologically-driven, less-principled staff; increased barriers to accessing high-quality journals via pay-walls; attacks on high-quality journals via DDOS; and over-reliance on non-peer-reviewed sources. The previously described routes could gradually, deliver significant shocks to a nation's healthcare capacity and gradually sink struggling facilities. This possibility is compounded through medical professional susceptibility to propaganda as evidenced by some within the anti-vaccination movement (Khazan, 2017).

## **6. Unsubstantiated specific drugs and medical equipment preferences**

A significant mode of the BCS-propaganda may be primed towards individuals most likely to be able to command large-scale contracts such as senior company officials or members of the government. Secondly, such may be targeted towards their influencers, such as lobbyists or sales-representatives. Tertiarily, propaganda may be aimed at lower workers and academicians. Following, such propaganda targeting can be aimed at civilians to influence upwards. With this hierarchy, the influence for the purchase and reception of monetary incentives for drugs is possible. This also applies to equipment as BCS-propaganda can affect the directing of equipment of ventilators or masks. In terms of COVID-19, it has been found that some masks can be less effective than no masks (*The mask matters: How masks affect airflow, protection effectiveness*, 2020). The consequences of propaganda that skews purchases and the use of ineffective equipment can be nearly as damaging as BCS propaganda on drugs and related material, and further, can distract from current, more effective equipment-based interventions. One more area worthy of investigation is the global divide between the Global North and South. COVID-19 has shown that a divide in access has existed as vaccines are readied, whereas the wealthier nations have expressed resistance in relaxing COVID-19 vaccine IP (Farge, 2020). While they are willing to share. This is where malicious actors could hypothetically push for wealthier nations to find means of extended leverage for increased access to resources or strengthen groups in weaker nations who may push back against vaccines. Although extreme, this illustrates not only the reach, but the dimensionality that notions of kickbacks for stakeholders can take on.

## **7. Potential for Industrial sabotage**

BCS propaganda affecting research could result in supply chain damage and to the adequate production of the products themselves. This can kickstart a chain reaction wherein a nation's long-term response to pandemic management is hindered. Between companies and communities, this can produce large-scale market imbalances. However, a further dark side to this is that of potential, substantial damage to literature and clinical trials. As a result, misinformation or disinformation entered could result in improper entries to journals, thus delaying or clouding solutions to other pathogens that rely on said data. For example, mRNA-based vaccines, although crafted shortly for COVID-19, relied on data from prior vaccines decades prior (Garde & Saltzman, 2020). In the future, it is possible that malicious actors may look to the sabotage of research to delay or stop progress decades down the line. There are also cross-field effects in the case of agriculture and pharmaceutical generation being entwined. For example, one can point to cases such as of tobacco plants being used to generate pharmaceutical or vaccine components. BCS-propaganda attacks on the use of such agricultural technology can affect both agricultural research and that of pharmaceutical research at the same time, not counting emergent research that results from innovations in either field (Palca, 2020). An example of a short-term benefit is exercised leverage, but long-term ramifications can be worldwide and cross-discipline.



## **8. Relevance to bioagents**

A conventional assumption about how a bioweapon should work is that it would be an exceptionally virulent agent that would decay outside the human body rather quickly when exposed to normal, atmospheric conditions. This makes sense in the context of using bioweapons as a tool for conventional warfare. An actor would not want to soften a target only to make it outright impossible for their own forces to take due to the risk of catching a disease of one's own manufacture. This would only be useful in a "Scorched Earth" strategy which most conventional conflicts tend to avoid due to the widespread environmental and economic damage as shown with experiments with anthrax, which limited use of the land due to threat perseverance (Hulme, 2011; Li, 2005; Wilson, 2005). However, if one were to make a biological weapon and optimize it for psychological damage as a backdrop to an information-warfare campaign, it makes more sense to create an agent that would last a relatively long time, even if its effects would not be particularly deadly, or its onset particularly acute. Prior to COVID-19, the suspected agent for such an attack could have been something like a weaponized strain of Tularemia (Michaelis, 1991; Sjöstedt, 2001). The longevity of such an attack would be useful for two reasons. The first would be through draining resources caused by complicated triage of the infected. The second would be through draining personnel resources from a hospital system, especially those understaffed.

This psychological edge would be lost if perpetrator-identification were easy. Logically, giving face to an external cause of a devastating disease would likely fast-track adversary mobilization. Therefore, utilizing a "quieter" disease and creating a fog of war, especially in places like cyberspace where the flow of information is either unchecked or the flow mechanism faces no consequences for displaying unverified information. This in turn incentivizes subtle weakening techniques. Following, distribution of propaganda becomes a necessity for the optimal use of certain strains of bioweapons, and this then reinforces the importance of health education as a modern defense foundation component.

## **9. How BCS can be affected through propaganda**

To discuss theoretical fog of war creation, prior insight on misinformation and disinformation through cyber networks, specifically through social media, are helpful to examine. For example, Facebook has long been a hotbed and topic for propaganda, but criticism reached new heights in 2016, wherein flooding by malicious actors resulted in meaningful propaganda injections which have been alleged as critical election factors in 2016 and beyond (Benkler, 2018; Bourassa et al, 2017; Faris et al, 2017; Fisher, 2020; Golovchenko, 2020; Howard and Kollanyi, 2016; Howard et al, 2018; Stewart et al, 2018; Walter and Jamieson, 2020; Woolley and Guilbeault, 2017). Relevant to Biology is that of Covid-19 vaccinations, of which inadequate to borderline acceptable public support has concerned healthcare workers and officials, triggering questions of what social media giants can do fix this (Iboi et al, 2020; Neergaard et al, 2020; Wilson and Wiysonge, 2020). The social-cyber engineering barrier combined with physical ramifications places this squarely within BCS, and social media companies wield heavy influence here as they support a marketplace and often gear, which both eliminates the typical "distance" a user has from another, via from a cyber interface and a physical interface (phones, VR, and other equipment). Considerable research especially notes that the internet and VR platforms can play a role in affecting behavior, and this is important considering the value of cyber-competency in leaders, even to military personnel (Day, 2020; Fuenekes et al, 2020; Sundararaj and Rejeesh, 2021; Kavanagh, 2020; Adžgauskaitė and Presavento, 2020; Kim et al, 2020; Machulska et al, 2020). How social media will ultimately guide BCS-evolution since it can be a conduit for health information, is important for reflection.

## **10. Other intersections and cost comparisons**

BCS-based propaganda can stretch to military personnel maintenance, but also shares an intersection with biologics research, in terms of military advantage with respect to concerns of Covid-19 research espionage (Lallie et al, 2020; Lee and Haupt, 2020). This intersection is additionally valuable given the presence of biodata and value of exploiting it as an alternative means of conflict for nations looking to make themselves more competitive at reduced expense. For example, \$10,000 affords few effective munitions, but can buy considerable ad-space and time through social media with the possibility of allowing the Global south to spread inaccurate news. As another example, at the time of this writing, one site listed costs per click on a platform at under \$0.40 (Birk, 2020; Lua, (n.d.). Compare with the cost of an M1A2 tank, MQ-1 Predator UAV, and M4 Carbine Rifle) at over 6 million, 4 million, and 700 USD, respectively (*ABRAMS TANK (M1A2)*; Curtis, 2012; *UNITED STATES AIR FORCE FY 2011 Budget Estimates 2010*). Cost comparisons are convincing for slight military budget restructuring, in innocuous ways; comparisons of traditional weapons to cyber weapons, for example, are already noted in the literature (Bates, 2020). Long-term consequences may of budget modeling and diversification can be manifold

across multiple sectors. For example, adapted to agriculture, this can take form in the disruption of meaningful legislation for GMO crops for a rival nation or on legislation that affects farmer-to-distributor transactions (Shukla et al, 2018; Botha et al, 2020; Oloo et al, 2020; Mashal et al, 2021). A further application of this may see propaganda aimed at influencing divestment from weapons programs to save other sectors between rivals.

BCS-propaganda could also be used to change course and depress containment efforts. Super-spreader events possess this capacity, and it is possible for malicious actors to tweak propaganda to trigger more (Lemieux et al., 2020; Fontes, Reyes, Ahmed, & Kinzel, 2020; Wells, 2020). This falls into the hands of repressive efforts by some authoritarian governments, which have promoted for curtailing abuse of speech popularization risks for long-term threats regarding free-speech protections among weakened yet freer governments. This noted, research on curtailing speech abuses without repressive cuts on speech may be favorable for the preservation of Western values. This is just one of many routes of societal change that could use additional investigation as pandemics impact politics and government function.

## 11. Conclusion

An individual's choices will be largely determined by the information they are exposed to. In the case of BCS threats, this information must be timely, verifiable, or demonstrably correct to make high-quality decisions. The primary concern as discussed above is that of BCS threats and the different stakeholder groups acquired data through different modalities. Each of those modalities can be violated in different ways with misinformation and disinformation. In particular, average citizens who utilize social media for their information may have platforms that lack information verification standards. Medical and industrial professionals tend to have higher standards, yet those protocols can also be violated through most labor-intensive means. Additionally, the funding of a BCS threat based on primarily disinformation is cost-effective when compared to conventional military munitions, programs, and operations.

## References

- Abrams, T. (M1A2). (n.d.). <https://web.archive.org/web/20131103110937/http://www.globalsecurity.org/military/library/budget/fy1999/dote/army/99m1a2.html>
- Adžgauskaitė, M., Abhari, K., & Pesavento, M. (2020, July). How Virtual Reality Is Changing the Future of Learning in K-12 and Beyond. In *International Conference on Human-Computer Interaction* (pp. 279-298). Springer, Cham.
- Aro, J. (2016). The cyberspace war: propaganda and trolling as warfare tools. *European view*, 15(1), 121-132.
- Bard, M. T. (2017). Propaganda, persuasion, or journalism? Fox News' prime-time coverage of health-care reform in 2009 and 2014. *Electronic News*, 11(2), 100-118. <https://doi.org/10.1177/1931243117710278>
- Bates, N. (2020). (publication). Comparing Cyber Weapons to Traditional Weapons Through the Lens of Business Strategy Frameworks. Egham, United Kingdom: Royal Holloway, University of London.
- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press. DOI:10.1093/oso/9780190923624.001.0001
- Berlivet, L., & Löwy, I. (2020). Hydroxychloroquine Controversies: Clinical Trials, Epistemology, and the Democratization of Science. *Medical anthropology quarterly*, 34(4), 525-541. <https://anthrosource.onlinelibrary.wiley.com/doi/full/10.1111/maq.12622>
- Birk, M. (2020, November 17). Understanding Facebook Ads Cost: 2019 & 2020 Benchmarks. <https://adespresso.com/blog/facebook-ads-cost/>
- Botha, A. M., Kunert, K. J., Maling'a, J., & Foyer, C. H. (2020). Defining biotechnological solutions for insect control in sub-Saharan Africa. *Food and Energy Security*, 9(1), e191.
- Broadcasting False Information. (2021, January 08). <https://www.fcc.gov/consumers/guides/broadcasting-false-information>
- Cavalcanti, A. B., Zampieri, F. G., Rosa, R. G., Azevedo, L. C., Veiga, V. C., Avezum, A., ... & Berwanger, O. (2020). Hydroxychloroquine with or without Azithromycin in Mild-to-Moderate Covid-19. *New England Journal of Medicine*, 383(21), 2041-2052.
- Chacin, A., & Klas, M. (2020, June 12). Reluctantly, under pressure, Florida disclosed COVID-19 data. What the numbers tell us. <https://www.miamiherald.com/news/coronavirus/article243195161.html>
- Coronavirus lockdown protest: What's behind the US demonstrations? (2020, April 21). <https://www.bbc.com/news/world-us-canada-52359100>
- Cox, D. (2021, January 11). How to stop the next pandemic. <https://www.wired.co.uk/article/next-pandemic>
- Curtis, R. (2012, April 20). U.S. Army places order for 24,000 M4A1 carbines with Remington. <https://web.archive.org/web/20120624094006/http://militarytimes.com/blogs/gearscout/2012/04/20/us-army-places-order-for-24000-m4-carbines-with-remington/>
- Day, A. E. (2020). Leading air force cyber warriors: Cyber wing commander competencies (Order No. 27831938). Available from ProQuest Dissertations & Theses Global. (2394300291).

- <http://proxy.lib.odu.edu/login?url=https://www.proquest.com/dissertations-theses/leading-air-force-cyber-warriors-wing-commander/docview/2394300291/se-2?accountid=12967>
- Ellyatt, H. (2020, November 17). No-lockdown Sweden toughens up restrictions as coronavirus cases rise. <https://www.cnn.com/2020/11/17/sweden-toughens-up-coronavirus-rules-as-infections-and-deaths-rise.html>
- Falcão, M. B., de Goes Cavalcanti, L. P., Filgueiras Filho, N. M., & de Brito, C. A. A. (2020). Case report: hepatotoxicity associated with the use of hydroxychloroquine in a patient with COVID-19. *The American journal of tropical medicine and hygiene*, 102(6), 1214-1216.
- Farge, E. (2020, November 20). Wealthy countries block COVID-19 drugs rights waiver at WTO - sources. <https://uk.mobile.reuters.com/article/amp/idUKKBN28020X>
- Feunekes, G. I., Hermans, R. C., & Vis, J. (2020). Public health nutrition communication in the Netherlands: From information provision to behavior change. *Handbook of Eating and Drinking: Interdisciplinary Perspectives*, 617-639.
- Fiedler, Beth Ann. Diffusion of COVID-19 in the United States: Politics, Social Determinants, or Neither? *Environmental Epidemiology*.2020.Preprint. DOI:10.13140/RG.2.2.22073.39523.
- Fisher, A. (2020). Demonizing the enemy: the influence of Russian state-sponsored media on American audiences. *Post-Soviet Affairs*, 1-16.
- Fontes, D., Reyes, J., Ahmed, K., & Kinzel, M. (2020). A study of fluid dynamics and human physiology factors driving droplet dispersion from a human sneeze. *Physics of Fluids*, 32(11), 111904. doi:10.1063/5.0032006
- Forman, R., Atun, R., McKee, M., & Mossialos, E. (2020). 12 Lessons learned from the management of the coronavirus pandemic. *Health Policy*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7227502/>
- Freckelton QC, I. (2020). Covid-19: Fear, quackery, false representations and the law. *International Journal of Law and Psychiatry*, 72, 101611. doi:10.1016/j.ijlp.2020.101611
- Garde, D., & Saltzman, J. (2020, November 10). The story of mRNA: From a loose idea to a tool that may help curb Covid. <https://www.statnews.com/2020/11/10/the-story-of-mrna-how-a-once-dismissed-idea-became-a-leading-technology-in-the-covid-vaccine-race/>
- Golovchenko, Y., Buntain, C., Eady, G., Brown, M. A., & Tucker, J. A. (2020). Cross-Platform State Propaganda: Russian Trolls on Twitter and YouTube During the 2016 US Presidential Election. *The International Journal of Press/Politics*, 1940161220912682.
- Gover, A. R., Harper, S. B., & Langton, L. (2020). Anti-Asian hate crime during the COVID-19 pandemic: Exploring the reproduction of inequality. *American journal of criminal justice*, 45(4), 647-667. <https://link.springer.com/article/10.1007/s12103-020-09545-1>
- Howard, P. N., & Kollanyi, B. (2016). Bots, #StrongerIn, and #Brexit: computational propaganda during the UK-EU referendum. *Available at SSRN 2798311*. DOI:10.1093/oso/9780190931407.001.0001
- Howard, P. N., Kollanyi, B., Bradshaw, S., & Neudert, L. M. (2018). Social media, news and political information during the US election: Was polarizing content concentrated in swing states?. *arXiv preprint arXiv:1802.03573*.
- Hulme, P. E. (2011). Biosecurity and the politics of fear. *Science*, 334(6053), 176-177.
- Iboi, E. A., Ngonghala, C. N., & Gumel, A. B. (2020). Will an imperfect vaccine curtail the COVID-19 pandemic in the US?. *Infectious Disease Modelling*, 5, 510-524. DOI:10.1016/j.idm.2020.07.006
- Kavanagh, D. J. (2020). 29 Changing Behavior in the Digital Age. *The Handbook of Behavior Change*, 416.
- Khazan, O. (2017, January 18). The shadow network of Anti-vax doctors. <https://www.theatlantic.com/health/archive/2017/01/when-the-doctor-is-a-vaccine-skeptic/513383/>
- Kim, M. J., Lee, C. K., & Jung, T. (2020). Exploring consumer behavior in virtual reality tourism using an extended stimulus-organism-response model. *Journal of Travel Research*, 59(1), 69-89.
- Lallie, H. S., Shepherd, L. A., Nurse, J. R., Erola, A., Epiphaniou, G., Maple, C., & Bellekens, X. (2020). Cyber security in the age of covid-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. *arXiv preprint arXiv:2006.11929*.
- Langin, K. 2. (2018, March 13). Fake news spreads faster than true news on Twitter-thanks to people, not bots. <https://www.sciencemag.org/news/2018/03/fake-news-spreads-faster-true-news-twitter-thanks-people-not-bots>
- Larson, C., Ghosal, A., & De Sousa, M. (2020, December 14). Scientists Focus on Bats for Clues to Prevent Next Pandemic. <https://www.nbcnewyork.com/news/national-international/scientists-focus-on-bats-for-clues-to-prevent-next-pandemic/2778773/>
- Lee, J. J., & Haupt, J. P. (2020). Scientific collaboration on COVID-19 amidst geopolitical tensions between the US 31. and China. *The Journal of Higher Education*, 1-27. 10.21203/rs.3.rs-37599/v2
- Lemieux, J. E., Siddle, K. J., Shaw, B. M., Loreth, C., Schaffner, S. F., Gladden-Young, A., . . . Macinnis, B. L. (2020). Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science*. doi:10.1126/science.abe3261
- Li, X. (2005). Blood-weeping Accusations: Records of Anthrax Victims. *Beijing: Zhong yang wen xian chu ban she*.
- Lua, A. (n.d.). Facebook Ads Cost: The Complete Guide to the Cost of Facebook Ads. <https://buffer.com/library/facebook-advertising-cost/>
- Machulska, A., Eiler, T. J., Grünewald, A., Brück, R., Jahn, K., Niehaves, B., ... & Klucken, T. (2020). Promoting smoking abstinence in smokers willing to quit smoking through virtual reality-approach bias retraining: a study protocol for a randomized controlled trial. *Trials*, 21(1), 1-10.
- Martins-Filho, P., Carvalho, A. C., de Melo, E. M., Mendes, M. L. T., & Santos, V. (2020). The "unbridled race" for using chloroquine and hydroxychloroquine to prevent or treat COVID-19 leads to shortages for patients with chronic

- inflammatory conditions and malaria in Brazil. <https://www.cmai.ca/content/unbridled-race-using-chloroquine-and-hydroxychloroquine-prevent-or-treat-covid-19-leads>
- Michaelis, A. R. (1991). Environmental Warfare. <https://www.tandfonline.com/doi/pdf/10.1179/isr.1991.16.2.97>
- Misra, S., Le, P. D., Goldmann, E., & Yang, L. H. (2020). Psychological impact of anti-Asian stigma due to the COVID-19 pandemic: A call for research, practice, and policy responses. *Psychological Trauma: Theory, Research, Practice, and Policy*. panel <https://pubmed.ncbi.nlm.nih.gov/32525390/DOI:10.1037/tra0000821>
- Neergaard, L., & Fingerhut, H. (2020). AP-NORC poll: Half of Americans would get a COVID-19 vaccine. *Associated Press*. <https://apnews.com/article/ap-norc-poll-us-half-want-vaccine-shots-4d98dbfc0a64d60d52ac84c3065dac55>
- Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., ... & Agha, R. (2020). The socio-economic implications of the coronavirus and COVID-19 pandemic: a review. *International journal of surgery*. DOI:10.1016/j.ijsu.2020.04.018
- Nivette, A., Ribeaud, D., Murray, A., Steinhoff, A., Bechtiger, L., Hepp, U., ... & Eisner, M. (2020). Non-compliance with COVID-19-related public health measures among young adults in Switzerland: Insights from a longitudinal cohort study. *Social Science & Medicine*, 268, 113370. <https://dx.doi.org/10.1016%2Fj.socscimed.2020.113370>
- Nordling, L. (2020). Unproven herbal remedy against COVID-19 could fuel drug-resistant malaria, scientists warn. *Science*. <https://www.sciencemag.org/news/2020/05/unproven-herbal-remedy-against-covid-19-could-fuel-drug-resistant-malaria-scientists>
- Oloo, B., Maredia, K., & Mbabazi, R. (2020). Advancing adoption of genetically modified crops as food and feed in Africa: The case of Kenya. *African Journal of Biotechnology*, 19(10), 694-701.
- Palca, J. (2020, October 15). Tobacco Plants Contribute Key Ingredient For COVID-19 Vaccine. <https://www.npr.org/sections/health-shots/2020/10/15/923210562/tobacco-plants-contribute-key-ingredient-for-covid-19-vaccine>
- Palmeira, V. A., Costa, L. B., Perez, L. G., Ribeiro, V. T., & Lanza, K. (2020). Do we have enough evidence to use chloroquine/hydroxychloroquine as a public health panacea for COVID-19?. *Clinics*, 75. <https://www.scielo.br/pdf/clin/v75/1807-5932-clin-75-e1928.pdf>
- Palmer, X., & Potter, L. (2021, February 6). *Commentary on Cataloguing Biological Assets and Academic DDoSing: Thoughts on Biocybersecurity in the Global South*. [Unpublished Presentation] presented at the Blacks in Cybersecurity Winter Conference 2021: Biohacking Village, [Online].
- Panagiotou, O. A., Contopoulos-Ioannidis, D. G., & Ioannidis, J. P. (2013). Comparative effect sizes in randomised trials from less developed and more developed countries: Meta-epidemiological assessment. *Bmj*, 346(Feb12 1). doi:10.1136/bmj.f707
- Pennisi, E. (2020, February 07). Scientists discover virus with no recognizable genes. <https://www.sciencemag.org/news/2020/02/scientists-discover-virus-no-recognizable-genes>
- Pushing for a people's vaccine for COVID-19. <https://www.doctorswithoutborders.org/what-we-do/news-stories/story/pushing-peoples-vaccine-covid-19>
- Reihani, H., Ghassemi, M., Mazer-Amirshahi, M., Aljohani, B., & Pourmand, A. (2021). Non-evidenced based treatment: An unintended cause of morbidity and mortality related to COVID-19. *The American journal of emergency medicine*, 39, 221-222. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7202810/>
- Richards, R. (2020). Evidence on the accuracy of the number of reported Covid-19 infections and deaths in Lower-Middle Income countries. <https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/15576>
- Robbins, R. (2017, September 10). Most Experimental Drugs are Tested Offshore—Raising Concerns about Data. <https://www.scientificamerican.com/article/most-experimental-drugs-are-tested-offshore-raising-concerns-about-data/>
- Scheid, J. L., Lupien, S. P., Ford, G. S., & West, S. L. (2020). Commentary: physiological and psychological impact of face mask usage during the COVID-19 pandemic. *International journal of environmental research and public health*, 17(18), 6655. <https://www.mdpi.com/1660-4601/17/18/6655/htm>
- Scientists focus on bats for clues to prevent next pandemic. (2020, December 14). <https://www.modernhealthcare.com/safety-quality/scientists-focus-bats-clues-prevent-next-pandemic>
- Section 230 of the Communications Decency Act. (n.d.). <https://www.eff.org/issues/cda230>
- Shukla, M., Al-Busaidi, K. T., Trivedi, M., & Tiwari, R. K. (2018). Status of research, regulations and challenges for genetically modified crops in India. *GM crops & food*, 9(4), 173-188.
- Sjöstedt, A. (2007). Tularemia: history, epidemiology, pathogen physiology, and clinical manifestations. *Annals of the New York Academy of Sciences*, 1105(1), 1-29. Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election. *Berkman Klein Center Research Publication*, 6.
- Stewart, L. G., Arif, A., & Starbird, K. (2018, February). Examining trolls and polarization with a retweet network. In *Proc. ACM WSDM, workshop on misinformation and misbehavior mining on the web*.
- Sundararaj, V., & Rejeesh, M. R. (2021). A detailed behavioral analysis on consumer and customer changing behavior with respect to social networking sites. *Journal of Retailing and Consumer Services*, 58, 102190.
- Taylor, A. (2020, September 25). As debate over Sweden's covid-19 response continues, cases in the country are on the rise. <https://www.washingtonpost.com/world/2020/09/25/sweden-coronavirus-debate-lockdown-new-cases/>
- The mask matters: How masks affect airflow, protection effectiveness. (2020, December 15). [https://www.eurekalert.org/pub\\_releases/2020-12/aiop-tmm121020.php](https://www.eurekalert.org/pub_releases/2020-12/aiop-tmm121020.php)

**Xavier-Lewis Palmer, Ernestine Powell and Lucas Potter**

- Thrasher, S. (2020, December 01). World AIDS Day Is a Grim Reminder That We Have Many Pandemics Going On. <https://www.scientificamerican.com/article/world-aids-day-is-a-grim-reminder-that-we-have-many-pandemics-going-on/>
- UNITED STATES AIR FORCE FY 2011 Budget Estimates [PDF]. (2010, February). United States Air Force. <https://web.archive.org/web/20120304052331/http://www.saffm.hq.af.mil/shared/media/document/AFD-100128-072.pdf>
- Wallace-Wells, D. (2020, December 07). We Had the Vaccine the Whole Time. <https://nymag.com/intelligencer/2020/12/moderna-covid-19-vaccine-design.html>
- Walter, D., Ophir, Y., & Jamieson, K. H. (2020). Russian Twitter Accounts and the Partisan Polarization of Vaccine Discourse, 2015–2017. *American Journal of Public Health*, 110(5), 718-724.
- Wells, R. (2020, November 19). Researchers identify features that could make someone a virus super-spreader. <https://phys.org/news/2020-11-features-virus-super-spreader.html>
- Wilson, B., & Gunderson, P. (2005). Biological and chemical terrorism and the agricultural health and safety community. [https://doi.org/10.1300/J096v10n02\\_02](https://doi.org/10.1300/J096v10n02_02)
- Wilson, S. L., & Wiysonge, C. (2020). Social media and vaccine hesitancy. *BMJ Global Health*, 5(10), e004206.
- Woolley, S. C., & Guilbeault, D. (2017). Computational propaganda in the United States of America: Manufacturing consensus online. <https://blogs.oii.ox.ac.uk/politicalbots/wp-content/uploads/sites/89/2017/06/Comprop-USA.pdf>
- Woolley, S. C., & Howard, P. N. (Eds.). (2018). *Computational propaganda: political parties, politicians, and political manipulation on social media*. Oxford University Press.
- Ziems, C., He, B., Soni, S., & Kumar, S. (2020). Racism is a virus: Anti-asian hate and counterhate in social media during the covid-19 crisis. *arXiv preprint arXiv:2005.12423*.

# Bio-Cyber Operations Inspired by the Human Immune System

Seyedali Pourmoafi and Stilianos Vidalis

University of Hertfordshire, Hatfield, Hertfordshire

[S.pourmoafi@herts.ac.uk](mailto:S.pourmoafi@herts.ac.uk)

[S.vidalis@herts.ac.uk](mailto:S.vidalis@herts.ac.uk)

DOI: 10.34190/EWS.21.089

**Abstract:** Bio-Cyber operation is a new field of research that is inspired by the Human Immune System. The human body has found solutions for problems that cybersecurity professionals have been trying to resolve for the past few decades. Cybersecurity should draw lessons from the human immune system on how to detect and deter attacks. Systems and devices are likely to leak sensitive information or data. A 'cyber immune' technology can be used to detect unknown cyber-attacks and provide a powerful mechanism for defence. In this paper we focus on work that describes the recent advances on Bio-Cyber operations, and we present our conceptual cyber operations model. By looking into the field of human biology we aspire to provide significant insight into the bio-cybersecurity domain.

**Keywords:** Bio-Cyber operation, human immune system, biological-inspired computing

---

## 1. Introduction

2020 has been an unprecedented time in our lifetime. Cybersecurity incidents are on the rise, with threat agents targeting crisis-related assets. We speculate that threat agents are taking advantage of new opportunities that have been created due to the changes in the operational framework of companies around the world. Threat agents are those individuals or groups of people that can manifest a threat (Vidalis and Jones, 2003).

The notion of threat can be defined as the function of a threat agent exploiting a vulnerability of an asset. An asset is a thing (or a service) that either now or in the future can generate or has the potential to generate (directly or indirectly) revenue for the business (Vidalis and Jones, 2003).

Today, our society is depended on the Global Information Environment (GIE). Cyber-attacks can happen instantly. Modern threats, for example zero-days and Advanced Persistent Threats (APT) poses very serious risk to the GIE. The Internet of Things (IoT), where cyber-physical systems cooperate and communicate with each other, presents devices and systems that usually use firmware or legacy code that was not improved with cyber threats in mind (Bhopi and Dongre, 2016).

As we continue to embrace and use all the benefits of the cyber operation, we need to move to a more secure world that focuses on behaviours within a network to identify normal behaviours from abnormal behaviours both at the group and individual level.

The aim of trying to "secure" the information is not new. In order to have a defence chance, cyber operation, just like the human body, must be defended through focusing and understanding the information infrastructure that is at risk at any one time. To serve this purpose, we need to start to implement a cyber immune system that acquires from its environment to avoid repeating attacks or problems and combat new ones (Kar, 2016).

In this paper we cut across scientific boundaries, discussing correlations between cybersecurity notions and concepts with biological notions and concepts. We draw inspiration from the human immune system, and we present a conceptual model for cyber operations.

## 2. The global information environment

The aforementioned concepts are not new. What is new, and what is constantly changing, is the environment into which we, as individuals, and of course businesses have to operate (Schwartz, 2018). The US DoD has defined the Information Environment as "the aggregate of individuals, organizations and systems that collect, process disseminate or act on information (Parn and Edwards, 2019). The global information environment has some characteristics that uniquely describe it. It is by far a non-static environment that is making use of change as a catalyst in order to address business needs (client needs) for constantly generating revenue streams. It can consist of a plethora of devices, using diverse mobile architectures. Furthermore, it is hyper-charged in that

every device is performing several roles, offering services to other devices, all related to strangeness (Truong, Diep and Zelinka, 2020).

Due to technological advances, communication is being exchanged at unprecedented rates. 2.5 quintillion bytes of data are created each day globally. The average decision maker is being presented with 34 GB data each day. The modern information environment attempts (and not with great success) to constantly change in order to enable for the aforementioned communication to take place (Abbas Ahmed, 2016).

Any member (or operator) of the environment can produce and consume information, often at the same time. The average decision maker today has access to platforms allowing for capabilities that until the recent past were limited to specific sections of the defense community (Carley and Cervone, 2018). In addition, everyone wants to get involved, but more of the point, there is the expectation that everyone will be allowed to be involved in every decision originating from the environment, and for the environment.

Finally, almost everyone listens. It is hard to define boundaries (internal or external) to the information environments. Ownership and responsibilities can be two grey areas.

The Information Environment is not the only concept that has become global in the current computing evolution era. Cyber Operators have also become global (Parn and Edwards, 2019). Instead of looking at local systems and local environments, cyber operators are now truly operating across geographical boundaries, cutting across jurisdictions, having to manage and protect complex interrelationships between tangible and intangible assets (Von Solms and Van Niekerk, 2013). We argue that this also applies to cyber as well as kinetic operations.

### **3. The challenge of the GIE**

The same concepts that were discussed as attributes of the GIE in a previous section, can also be seen as challenges that global cyber operators are trying to overcome. The main objective of any business is the generation of revenue. From the perspective of a cyber operator, an increase in revenue will be the by-product of achieving information superiority for their organization. In agreement to (Lin, 2019), information superiority is the operational advantage derived from the ability to collect, process and disseminate an uninterrupted flow of information, while exploiting or denying an adversary to do the same. Information superiority is a state achieved as the result of successful information operations (Vishnevsky, Kozyrev and Larionov, 2014)

Information operations are continuous acts of force in the GIE (Parn and Edwards, 2019), to compel our adversaries to do our will. We classify information operations in the following types:

- Intelligence Operations (Inc counterintelligence): Cyber intelligence is a cyber discipline that exploits a wide range of information collection and analysis methods to provide decisions and direction to cyber operation units and cyber commanders (Eom, 2014).
- Psychological Operations: Psychological operations that transport information for instance broadcasting satellite radio messages, with the object of manipulating the views of organizations, foreign governments, or individuals (Hollis, 2007).
- Deception: Deception usually uses computer networks and information technology to intentionally deceive adversary decision-makers as friendly military abilities, deliberate and operation, so the adversary takes an action in ways that aid to friendly forces' mission (Hollis, 2007).
- Computer network operations (CNO) (CNA, CND, CNE)
- *Computer Network Attack (CNA)*
- *Computer Network Defense (CND)*
- *Computer Network Exploitation (CNE) (Whyte, 2016).*
- Situational Awareness Operations: Situational awareness is the understanding of environmental events and elements with regard to space or time, the perception of their meaning, and the plan of their future status (Seppänen and Virrantaus, 2015).

- **Operational Security:** Operational security (OPSEC) which is known as procedural security, is a risk management operation, that encourages managers to take in the process from the vision of an opponent in order to protect sensitive data and information from the leak into the wrong hand (Haddad *et al.*, 2011).
- **Information Security:** Information security is a set of activities that try to keep data and information secure from unauthorized alterations or access (Von Solms and Van Niekerk, 2013).
- **Physical Security:** Physical security is the activity of personnel, data, networks, software, and hardware from physical action and event that could reason serious loss or damage to an agency, enterprise, or institution (Homes, 2018).

It is not our intension to discuss the above operation types in isolation. Instead, we will discuss their correlations and how we have used the human immune system for modelling interrelationships between operations, aiming to better defending of an information environment

#### **4. The human immune system correlated to CyberOps**

Cyber Operation (CyberOps) is an interdisciplinary field encompassing the whole scope of cyberspace and related activities that are both technical and non-technical in nature, for instance ethical, legal, human-centered, etc. Cyber Operations is a supplementary subject to Cybersecurity. Cyber Operations places special emphasis on techniques and technologies applicable to all system and operational levels (Smeets, 2018).

If cyber operations are compared to the human body, then cyber-attacks can compare to viruses.

The human body has a significant effective immune mechanism called the immune system which can protect and detect wide range of harmful agents, such as microbes, viruses, and parasites which are known as pathogens (Nicholson, 2016). The human immune system includes special organs, cells, and chemicals that can fight pathogens or any infection. The main part of the immune system is made up of blood cells, antibodies, the complement system, the lymphatic system, the spleen, the thymus, and the bone marrow, they are an internal part of the immune system. However, the external part of our immune system is the skin, which fend off external threats similar to a firewall. It is continuously adaptive and renewed. Our immune system monitors the internal environment regularly and constantly (Parham, 2015).

In general, the immune system contains the adaptive immune system as well as an innate immune system. The vertebrates and invertebrates have innate immunity whereas the adaptive immune system is found only in invertebrates (Parham, 2015). If a pathogen breaks physical barriers of the body, the innate immune system presents an immediate response, but it is a non-specific response, and it is not able to confer long term immunity. The adaptive immune system can present a pathogen specific, tailored response. If the same pathogen enters the body, the response is remembered, and the immune system can present a quick and specific response to the antigen. The immune system has the ability to learn, to remember, and to identify patterns. Likewise, the adaptive immune system acts as the memory of the immune system (Nicholson, 2016). The adaptive immune system includes the lymphocytes which consist of the significant types of B-cells and T-cells. B cells recognize pathogens when antibodies on their outside connect to a certain foreign antigen. When an antigen enters the body, the immune system sends a signal to direct specific immune cells, known as killer T cells, to the infection's site. The killer T cells annihilate cells that are affected by viruses or any other pathogens and dysfunctional cells. An adaptive immune system recognizes and neutralizes antigens (think of ransomware or Trojans as antigens) by erasing or quarantine them. This function is similar to the biological systems' function which kills cells affected by known or unknown viruses (Okamoto and Tarao, 2016). In case B cells and T cells are activated, they reproduce, and some of their children become long-lived memory cells. A cyber immune system acts very similar to this behavior.

#### **5. Bio-inspired cyber security**

All living creatures take advantage of the wide range of natural forms of protection to help them to reduce as well as to avoid possible risks and adapt to their environment for survival. These biological predispositions and instincts, which are different among living creatures, can be synthetically implemented and replicated to the cyber immune systems in order to increase the system's resilience when facing an attack (Guthikonda *et al.*, 2017). This method of cybersecurity (bio-inspired cybersecurity) can be categorized by the weaknesses in security systems such as transparency, bioinformatic analyses, security standards, and cryptography (Wlodarczak, 2018).



## **5.1 Transparency**

The Ant-Based Cyber Defense (ABCD) (Fink *et al.*, 2014) is a bio-inspired method that monitors and defends the computer networks by adapting the algorithms and metrics by patterning the social insects, especially ants. ABCD such as a society in which humans and autonomous adaptive software agents cooperate (Fink *et al.*, 2014). In the changing attackers' strategies, ABCD can supply fast, stable adaptation and a dynamic environment. The ABCD software uses a biological concept known as swarm intelligence to implement. Swarm intelligence not only affected making of decisions but also allows ants to communicate and gathering in a risk's instance in a colony. ABCD utilizes a hierarchy of digital instruments, which are looking for harmful activity by comparing different machines in the system (Korczyński *et al.*, 2016).

## **5.2 Bioinformatic analyses**

Bioinformatic analyses transform cycles of behaviors or instructions to the data set for each same metrics that may be estimated by applying protein-sequencing tools (Sinha, 2014). Similarity can end up in the interrelationship groupings, description, and fast recognition of beforehand hidden suspect behaviors or attack designs. Similarity can end up in the interrelationship groupings, description, and fast recognition of beforehand hidden suspect behaviors or attack designs. Bioinformatic analyses are applied to an attack on a behavioral motif that has been seen before and to help determine how the malware can be neutralized (Ahsan, Gomes and Denton, 2018).

## **5.3 Cryptography**

Cryptography is the field of science that tries to keep information secure by changing it into a form that unintended recipients cannot recognize or understand (Kahate, 2008).

### *5.3.1 Chirp RF signals*

One of the first bio-inspired signals which often apply to the RF application is the chirp signal. The sharp signal inspires by the chirping sound, which made by birds (Zhang *et al.*, 2016). Fluctuating in chirps can provide multiple layers of information at the same time and different species of birds or social groups are not able to interpret the detail of the messages or undetectable by observers. By this capability, able birds to send messages to their destination receiver without understood by others (Korczyński *et al.*, 2016).

Radio Frequency (RF) Steganography is a way in which sensitive and secret data or messages hiding inside ordinary messages. Chirp radar signals are known as radar messages which modified with secret signals and hidden in a radio frequency waveform (Daras, 2018). Such minor changes can be recognized as a kind of embedded modulation that sends a sensitive or secure message which only understood by its target receiver. Linear chirp signals that emit from the radar are secured with a communication signal then receive and decoded in their intended receiver (Kar, 2016).

Chirp Radar signals are implemented essentially in monitoring technologies; however, they have also been used in military and covert processes because of their capability to efficiently defend and transmit sensitive and secret information and messages (Liu and Fok, 2021).

## **5.4 Security standards and metrics**

In general, hierarchy trees imitate a real tree structure. These structures are widely used, as they associate with computer science, they can be used to explain system processes where relationship paths exist (Guthikonda *et al.*, 2017).

### *5.4.1 Vulnerability tree models*

As figure 1 shows, vulnerability trees have modelled as hierarchy trees that the top hierarchy presents the main goal of the attacks (Aliyan *et al.*, 2020). Vulnerability trees can be exploited the top vulnerability, known as parent vulnerabilities (it has symbolised with a capital 'V'). There are numerous ways that such a top vulnerability can be exploited. Each of these ways have constituted a branch of the tree. Each of these ways can consider as one of the branches of the tree and leaves can constitute child vulnerabilities (they have symbolised with the lower case 'v') (Vidalis and Jones, 2003).

These models usually demonstrate the relationship between one vulnerability and other vulnerabilities and/or steps that a threat agent should carry out in order to achieve the top of the tree (Vidalis and Jones, 2003).

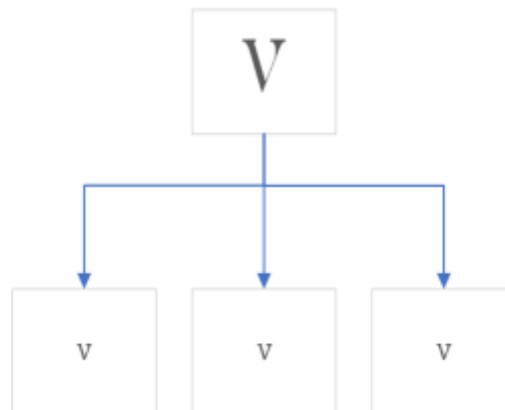


Figure 1: Vulnerability tree ("V" known as parent and "v" consider as a child)

#### 5.4.2 Attack tree models

In recent years, attack trees have been significantly developed to identify processes by which attackers or malicious users try to break or exploit computer software and/or network. An attack tree is a conceptual diagram indicating how a target, or an asset, might be attacked. Attack trees are extensively applied to threat scenarios in a concise and intuitive manner, which is suitable to express security information to non-experts (Kordy *et al.*, 2014). As figure 2 illustrates, which can be considered as multi-level diagrams and inspire the relationship between different parts of the real tree for example root, leaves, and children. The top of the tree is considered a top attack (Somestad, Ekstedt and Holm, 2013). From the bottom up, child nodes are the status that should be satisfied to construct the direct parent node true; when root is content the attacks are complete. These models are usually used to experiment with control centers and communication devices. In order to recognize and countermeasure against modelled viral and Internet attacks, they improve cybersecurity systems by producing flexibility in facing future attacks (Kordy *et al.*, 2014).

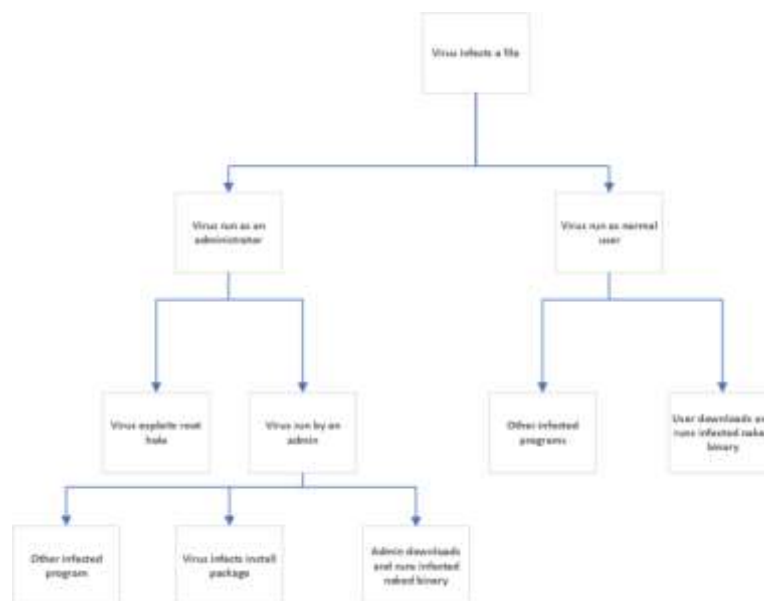


Figure 2. Attack tree model for virus

## 6. Cyber immune system

Cyber immunity is a bio-inspired method based on the human adaptive immune system that is able to learn and recognize attacks with unknown signatures (Igbe, 2019). Cybersecurity can be defined as a combination of technologies and processes designed and produced to protect and defend networks, computers, data, and programs from possible attack or risk, unauthorized access, variations, or any destruction.

The DNA of a virus changes continuously, so the immune system should adapt to detect the signature of the virus. Likewise, in cybersecurity, we face an ever-evolving adversary. Due to new attacks being unknown, it is very difficult to determine the signature of the attack from previous ones (Buczak and Guven, 2016).

Bio-inspired cyber immunity is able to learn and recognize the attacks with unknown signatures such as the human adaptive immune system (Figure 3) (Wlodarczak, 2018). A cyber immune system tries to learn what is the algorithm of the network traffic during a specific time, instead of learning attack signatures (Guthikonda *et al.*, 2017). Once learned, it determines the possibility that a certain abnormal algorithm is malicious. It regularly updates its information and results based on the new observation. It is able to interrupt an attacking agent by surveillance it and recognition what information the cause is and where it comes from. One of the main responsibilities in data mining is to find and select the monitoring points (Kar, 2016)

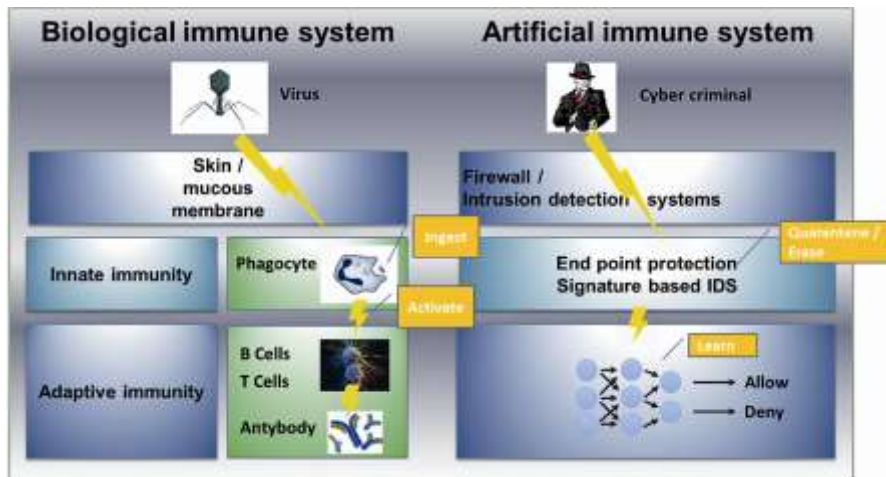


Figure 3: Brief comparison of biological and artificial immune system (Wlodarczak, 2018)

It should be noted that the immune system usually considers a metaphor. There is no connection between artificial immune systems and biological in the "mechanisms" of inherent and adaptive Immunity (Tarao and Okamoto, 2016).

### 7. Proposed research work

In this study, we have tried to show the relation between the biology and cyber security (Table 1. indicates some exaple of them). These similarities between biology and cyber operations have introduced the field of bio-inspired cybersecurity, and it has been recognized by cyber professionals. Professional recognition of overlap between biology and cyber operation has also emerging in the field of cybersecurity, which can identify blurring of the domain borders between biology and cyber and acknowledges new risks arising from the growingly cyber-physical nature of biotechnology. For example, today's, DNA scientists able to design genetic code that lets hackers inject malware into computer systems or devices. On the other hand, the accessibility of DNA sequences in bioinformatics systems suggests the probability for hackers to change genetic code in a way that converts a benign organism into a harmful pathogen (Peccoud *et al.*, 2018).

Table 1: The link between biology and cyber-security (HASSAN, MYLONAS and Vidalis, 2016).

Function	IT Infrastructure Action or Term	Cell Biology Action or Term
Barrier defense	Exterior Router Packet Filter/Stateful Inspection Firewalls Intrusion Detection Systems	Plasma Membrane/ Plant cell wall Oligasaccharins "oxidative brust"
Barrier Transmission and Communication	Tunneling protocols Secure Sockets Layer Virtual Private Networks Advanced Encryption Mechanisms Network Ports	Variety of Member Channels Gap Junctions Facilitated Diffusion and Transporters (i.e. Glucose) Extrasellar matrix signaling
Internal Organization	Internal Firewalls Network DMZ (Buffer Zones)	Membrane- bound organelles, mitochondria

Function	IT Infrastructure Action or Term	Cell Biology Action or Term
		Nuclear pore complex, double membrane envelope
Internal Routing and Sorting	Email, standard Fax, telephone IPv6 Routers, routing Table	Endocytosis, Exocytosis Golgi Apparatus Cell Nucleus
Virus Defense and Response	Infection Carrier System Cleansing (anti-virus s/w) System Isolation System Corruption	Infection Carrier Intracellular digestion, endocytosis Cell Division Cell Death

## 8. Conclusions

In today's world cyber-intrusions has increased significantly in computer systems, these figures rise every year. This so-called technological revolution may cause digital technology to become more complicated, and with it, cyber-attacks, in the near future. This is mainly because the structure of our world revolves around modern technology and digital media, so it is necessary to detect attacks and vulnerabilities with the proper security countermeasure, and bio-inspired cybersecurity can pave the way to achieve this goal. The body's self-defence mechanism is one of the best phenomena of biology. We should consider the human body as a great example of how modern systems must adapt to defeat attacks or threats.

In the future, bio-cyber security can be improved with innovation in genetic engineering and biomedical science.

## References

- Abbas Ahmed, R. K. (2016) 'Security Metrics and the Risks: An Overview', *International Journal of Computer Trends and Technology*, 41(2), pp. 106–112. doi: 10.14445/22312803/ijctt-v41p119.
- Ahsan, M., Gomes, R. and Denton, A. (2018) 'SMOTE Implementation on Phishing Data to Enhance Cybersecurity', *IEEE International Conference on Electro Information Technology*. IEEE, 2018-May, pp. 531–536. doi: 10.1109/EIT.2018.8500086.
- Aliyan, E. et al. (2020) 'Decision tree analysis to identify harmful contingencies and estimate blackout indices for predicting system vulnerability', 178(August 2019). doi: 10.1016/j.epr.2019.106036.
- Bhopi, S. K. and Dongre, N. M. (2016) 'Study of Dynamic Defense technique to overcome drawbacks of moving target defense', *Proceedings - IEEE International Conference on Information Processing, ICIP 2015*. IEEE, pp. 637–641. doi: 10.1109/INFOP.2015.7489461.
- Buczak, A. L. and Guven, E. (2016) 'A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection', *IEEE Communications Surveys and Tutorials*. IEEE, 18(2), pp. 1153–1176. doi: 10.1109/COMST.2015.2494502.
- Carley, K. M. and Cervone, G. (2015) 'Social Cyber-Security', pp. 1–6.
- Daras, N. J. (2018) *Computation, Cryptography, and Network*.
- Eom, J. ho (2014) 'Roles and responsibilities of cyber intelligence for cyber operations in cyberspace', *International Journal of Software Engineering and its Applications*, 8(9), pp. 137–146. doi: 10.14257/ijseia.2014.8.9.11.
- Fink, G. A. et al. (2014) 'Defense on the move: Ant-based cyber defense', *IEEE Security and Privacy*. IEEE, 12(2), pp. 36–43. doi: 10.1109/MSP.2014.21.
- Guthikonda, A. et al. (2017) 'Bio-inspired innovations in cyber security', *2017 14th International Conference on Smart Cities: Improving Quality of Life Using ICT and IoT, HONET-ICT 2017*, 2017-January, pp. 105–109. doi: 10.1109/HONET.2017.8102212.
- Haddad, S. et al. (2011) 'Operational security assurance evaluation in open infrastructures', *6th International Conference on Risks and Security of Internet and Systems, CRISIS 2011*. IEEE, 15408. doi: 10.1109/CRISIS.2011.6061831.
- HASSAN, M., MYLONAS, A. and Vidalis, S. (2016) 'When Biology Meets Cybersecurity: A Review', *Journal of Information Systems Security*, 12(3), pp. 177–199.
- Hollis, D. B. (2007) 'No Title', 11.
- Homes, I. S. (2018) 'Cyber and Physical Security Vulnerability Assessment for IoT-Based Smart Homes', pp. 1–17. doi: 10.3390/s18030817.
- Igbe, O. (2019) 'Artificial Immune System Based Approach to Cyber Attack Detection', (January). Available at: <http://search.proquest.com/openview/b28ad74cae4ed6b71fb8a3ed7988b389/1?pq-origsite=gscholar&cbl=18750&diss=y>.
- Kahate, A. (2008) 'Cryptography-Network-Security-Atul-Kahate.Pdf', p. 535.

- Kar, A. K. (2016) 'Bio inspired computing - A review of algorithms and scope of applications', *Expert Systems with Applications*. Elsevier Ltd, 59, pp. 20–32. doi: 10.1016/j.eswa.2016.04.018.
- Korczynski, M. et al. (2016) 'Hive oversight for network intrusion early warning using DIAMOND: A bee-inspired method for fully distributed cyber defense', *IEEE Communications Magazine*, 54(6), pp. 60–67. doi: 10.1109/MCOM.2016.7497768.
- Kordy, B. et al. (2014) 'Attack-defense trees', *Journal of Logic and Computation*, 24(1), pp. 55–87. doi: 10.1093/logcom/exs029.
- Lin, H. (2019) 'The existential threat from cyber-enabled information warfare', *Bulletin of the Atomic Scientists*. Routledge, 75(4), pp. 187–196. doi: 10.1080/00963402.2019.1629574.
- Liu, Q. and Fok, M. P. (2021) 'Bio-inspired photonics – marine hatchetfish camouflage strategies for RF steganography', *Optics Express*, 29(2), p. 2587. doi: 10.1364/oe.414091.
- Nicholson, L. B. (2016) 'The immune system', *Essays in Biochemistry*, 60(3), pp. 275–301. doi: 10.1042/EBC20160017.
- Okamoto, T. and Terao, M. (2016) 'Toward an artificial immune server against cyber attacks', *Artificial Life and Robotics*. Springer Japan, 21(3), pp. 351–356. doi: 10.1007/s10015-016-0282-9.
- Parham, P. (2015) 'The Immune System'.
- Parn, E. A. and Edwards, D. (2019) 'Cyber threats confronting the digital built environment: Common data environment vulnerabilities and block chain deterrence', *Engineering, Construction and Architectural Management*, 26(2), pp. 245–266. doi: 10.1108/ECAM-03-2018-0101.
- Peccoud, J. et al. (2018) 'Cyberbiosecurity: From Naive Trust to Risk Awareness', *Trends in Biotechnology*. Elsevier Ltd, 36(1), pp. 4–7. doi: 10.1016/j.tibtech.2017.10.012.
- Schwartz, N. A. (2005) 'Joint Publication 3-16', (August 2018). Available at: [http://www.bits.de/NRANEU/others/jp-doctrine/jp3\\_26\(05\).pdf](http://www.bits.de/NRANEU/others/jp-doctrine/jp3_26(05).pdf).
- Seppänen, H. and Virrantaus, K. (2015) 'Shared situational awareness and information quality in disaster management', *Safety Science*, 77, pp. 112–122. doi: 10.1016/j.ssci.2015.03.018.
- Sinha, S. (2014) 'Study on Agents Based Meta-Heuristic Approach for Cyber Security Defense Mechanism'.
- Smeets, M. (2018) 'The Strategic Promise of Offensive Cyber Operations', *Strategic Studies Quarterly - Fall*, 12(3), pp. 90–113.
- Sommestad, T., Ekstedt, M. and Holm, H. (2013) 'The cyber security modeling language: A tool for assessing the vulnerability of enterprise system architectures', *IEEE Systems Journal*. IEEE, 7(3), pp. 363–373. doi: 10.1109/JSYST.2012.2221853.
- System, A. B. C. D. (2018) 'Cyber Immunity - A Bio-Inspired Cyber Defense System', (November). doi: 10.1007/978-3-319-56154-7.
- Terao, M. and Okamoto, T. (2016) 'Toward an Artificial Immune Server against Cyber Attacks: Enhancement of Protection against DoS Attacks', *Procedia Computer Science*. The Author(s), 96, pp. 1137–1146. doi: 10.1016/j.procs.2016.08.156.
- Truong, T. C., Diep, Q. B. and Zelinka, I. (2020) 'Artificial intelligence in the cyber domain: Offense and defense', *Symmetry*, 12(3), pp. 1–24. doi: 10.3390/sym12030410.
- Vidalis, S. and Jones, A. (2003) 'Using vulnerability trees for decision making in threat assessment. School of Computing Technical Report CS-03-2', *Technical Report CS-03-2, School of Computing, University of Glamorgan*, (June), pp. 1–13. Available at: <http://books.google.com/books?hl=en&lr=&id=zrg3cMbSWjwC&oi=fnd&pg=PA329&dq=Using+Vulnerability+Trees+for+Decision+Making+in+Threat+Assessment&ots=1hYPnIj62q&sig=f2fWuoNAFJbsteuDB9wPDVAjby>.
- Vishnevsky, V., Kozyrev, D. and Larionov, A. (2014) *Distributed Computer and Communication Networks, Communications in Computer and Information Science*. doi: 10.1007/978-3-319-05209-0.
- Von Solms, R. and Van Niekerk, J. (2013) 'From information security to cyber security', *Computers and Security*. Elsevier Ltd, 38, pp. 97–102. doi: 10.1016/j.cose.2013.04.004.
- Whyte, C. (2016) 'Ending cyber coercion: Computer network attack, exploitation and the case of North Korea', *Comparative Strategy*. Taylor & Francis, 35(2), pp. 93–102. doi: 10.1080/01495933.2016.1176453.
- Zhang, Z. et al. (2016) 'Bio-inspired RF steganography via linear chirp radar signals', *IEEE Communications Magazine*. IEEE, 54(6), pp. 82–86. doi: 10.1109/MCOM.2016.7497771.

# Space Cyber Threats and Need for Enhanced Resilience of Space Assets

Jakub Pražák

Faculty of Social Sciences, Charles University, Prague, Czechia

[prazak.jakub94@gmail.com](mailto:prazak.jakub94@gmail.com)

DOI: 10.34190/EWS.21.006

**Abstract:** Space systems represent a vital part of great powers' critical infrastructure and are the concern of national security. Space provides essential civilian and military services which disruption would result in severe consequences leading from economic losses to catastrophic events. However, the emergence of "New Space" with an increased number of commercial enterprises and space systems, accompanied by the worsening relations between major space powers, raises severe concerns about space security and protection of space assets. In addition to that, most space systems are inherently dual-use; being a potential subject of both civilian and military utilization and interest. Accordingly, space powers are actively engaged in developing advanced counterspace capabilities that could be used to their advantage. Nevertheless, despite the security risks, little focus is paid to protecting space assets against cyber-attacks and security breaches. Cyberattacks may diverge from theft or denial of information to control or destruction of satellite systems, their subcomponents, or supporting infrastructure. Cyber-attacks provide advantageous disruptive capability due to its limited attribution, range of effect, flexibility, accessibility and affordable cost. Moreover, space systems often lack substantial cyber resilience, and cyber-attacks can be combined with other counterspace capabilities such as electronic or kinetic anti-satellite weapons for decisive outcomes. Thus, the article aims to illuminate the opportunities and consequences of malicious space cyber operations and address cyber protection of space assets. The article elaborates on cyber threats that could be used for offensive and hybrid operations in outer space and discusses their implications for space relations and space warfare strategies. The emphasis is given on the space weaponization and perception of space as a new theatre of war in the reflection of the deficient space regime and the need for enhanced cyber resilience of space assets.

**Keywords:** space security, cyber security, satellite, cyber resilience, space weapon, space warfare

---

## 1. Introduction

*"All warfare is based on deception,"* states a notorious quote by Sun Tzu (2021), expressing the idea that the key to a successful military operation is a bluffed and dazzled enemy. Cyberspace is a domain which grants a unique opportunity to conduct covert operations with a wide scale for harmful activity in other domains. Besides, outer space is a "final frontier" of humankind that provides daily services with essential dependence; however, it lacks precise domain awareness and could be compromised by an enemy cyber entity. Space systems represent a vital part of great powers critical infrastructure and are the concern of national security. Space provides essential civilian and military services, the disruption of which would result in severe consequences, from economic losses to catastrophic events. However, the emergence of "New Space", with an increased number of commercial enterprises and space systems, accompanied by the worsening relations between major space powers, raises severe concerns about space security and protection of space assets. In addition to that, most of the space systems are inherently dual-use; being a potential subject of both civilian and military utilization and interest. Accordingly, space powers are actively engaged in developing advanced counterspace capabilities<sup>1</sup> that could be used to their advantage.

Nevertheless, despite the security risks, little focus is paid to protecting space assets against cyberattacks and security breaches. Cyberattacks may diverge from theft or denial of information to control or destruction of satellite systems, their subcomponents, or supporting infrastructure. Cyber-attacks provide advantageous disruptive capability due to their limited attribution, range of effect, flexibility, accessibility and affordable cost (Harrison et al, 2020, p. 5). Moreover, space systems often lack substantial cyber resilience and cyber-attacks can be combined with other counterspace capabilities such as electronic or kinetic anti-satellite weapons for decisive outcomes. Thus, the article aims to illuminate the opportunities and consequences of malicious space cyber operations and address cyber protection of space assets. The article elaborates on cyber threats that could be used for offensive and hybrid operations in outer space and discusses their implications for space relations and space warfare strategies. Emphasis is placed on the space weaponization and perception of space as a new theatre of war in the reflection of the deficient space arms control regime and the need for enhanced cyber resilience of space assets.

---

<sup>1</sup> Counterspace capabilities can be used to "to deceive, disrupt, deny, degrade, or destroy space systems" (Weeden and Samson 2020, p. ix)

## 2. Space warfare in the New Space environment

Nowadays, space security is confronted with two significant challenges – rising tensions between major space powers and the emergence of New Space with private enterprises (Dobos and Prazak 2019). Space systems provide a wide array of commercial, scientific and military applications. The significance and value of satellites, which ensure vital functions for nations states, should thus arguably be a subject of critical infrastructure protection framework (Georgescu et al, 2019, p. 21). The Limited Test Ban Treaty from 1963 prohibited nuclear tests in outer space (United Nations Office for Disarmament Affairs 2020) and Article IV of the Outer Space Treaty from 1967 banned placing of weapons of mass destructions in Earth orbits (Union of Concerned Scientists 2004). However, there is no legally binding instrument that would put restrictions on the proliferation of other kinds of space weapons. Thus, the states developing a wide scale of counterspace capabilities ranging from electronic warfare to energy and kinetic anti-satellite weapons (Weeden and Samson 2020). Moreover, space technology is inherently dual-use (Johnson-Freese 2007, p. 27), constituting an indistinct line between military and civilian-commercial systems. In this connection, commercial enterprises are visibly penetrating the space sector for profit. Nevertheless, a sharp increase of space systems, which can reach more than a hundred thousand in the following decade (Analytical Graphics, Inc. 2020), presents an issue to space security in terms of a rising number of space debris but at the same time also establishes opportunities for cyber breach of those systems.

Based on the thoughts of Mackinder's geopolitical Heartland theory, Everett Dolman argued that Earth's orbits provide further access to deep space with unlimited resources and must be thus denied to any potential adversaries to ensure control over both outer space and Earth (Dolman 1999, pp. 89-93). However, absolute control over Earth by a sole state actor would arguably change the dynamics of states' sovereignty, since the single country could exercise power over the others (Havercroft and Duvall 2009, pp. 42-58). As Kleinberg (2007, pp. 17-18) pointed out, space is a national "Center of Gravity" which provides a strategic advantage, and its exploitation is thus essential for all terrestrial operations.

Still, the thoughts about the sole space actor are hard to imagine with current developments and technology available. According to Bowen (2020, pp. 228-268), *"space warfare is waged to deny access to outer space and advantage of utilizing space technology"* and *"spacepower is uniquely infrastructural and connected to Earth"*. Space attacks should then target enemy military systems that provide crucial support such as navigation, early warning, and communication to disable long-range weapons and unleash confusion between the ground forces. This advantage could increase the chances of a successful terrestrial attack before enemy consolidation of ground forces and replacement of space systems. Hence, even though space dominance may be limited in terms of ultimate control, space cyberattack could be implemented into space warfare strategies to exploit its specific features that would surpass disadvantages of traditional kinetic weapons, the utilization of which in a conflict would be easily recognized and result in the creation of a substantial amount of space debris.

## 3. Space cyber threats

Cyberattacks target computer systems or data itself. Concerning space systems, a cyberattack can focus on antennas on satellites and ground stations, the landlines that connect ground stations to terrestrial networks, and the user terminals that connect to satellites (Harrison et al, 2020, pp. 4-5). Though cyberattacks generally require advanced technical knowledge, the attack itself is relatively cheap and cost-effective. Thus, the cyberattack may be conducted by both state and non-state actors and may be potentially contracted to private groups or individuals (Harrison et al, 2020, p. 5). The main motivations behind cyberattacks are finance, espionage, disruption, politics and retaliation. Though the top motivation of cyber-attackers is finance (Lourenço and Marinos 2020, pp. 12-13), in the context of geopolitical tensions and rivalry, other motivations and their combinations should not be neglected or underestimated.

Threats may embody diverse meanings; nevertheless, in international relations a threat can be defined as *"a situation in which one agent or group has either the capability or intention to inflict a negative consequence on another agent or group"* (Rousseau and Garcia-Retamero 2007, p. 745). Though there is no internationally recognized definition of space weapons, the Secure World Foundation classifies cyber threats between counterspace capabilities which can be used to *"to deceive, disrupt, deny, degrade, or destroy space systems"* (Weeden and Samson 2020, p. ix) and can be thus used for offensive space operations with destructive effect. Moreover, Prague Security Studies Institute included space cyber operations among space hybrid operations that are defined as *"intentional, temporary, mostly reversible, and often harmful space"*

*actions/activities specifically designed to exploit the links to other domains and conducted just below the threshold of requiring meaningful military or political retaliatory responses”* (Robinson et al, 2018, p. 3). Such a definition highlights their malicious properties for outer space applications; however, also pointing out their unclear identification as a “weapon”, especially regarding its difficult attribution and general reversibility (Robinson et al, 2018, p. 3). A cyberattack can have various impacts, e.g. data loss, widespread disruptions, or permanent loss of a satellite (Harrison et al, 2020, p. 5). Nevertheless, cyberattacks can be measured by “effects-based doctrine”, by evaluation of its consequences and caused damage (Robinson et al, 2018, p. 18). Thus, if the attack results in considerable damage, destruction or permanent malfunction of the satellite, it could be regarded as an “armed attack”, similarly to, for instance, kinetic anti-satellite weapons (ASAT). The corruption of cyberspace is gradually internationally recognized; for instance, a serious cyberattack on NATO can trigger Article 5 about collective defence (NATO 2019). However, the threshold of conflict in outer space may be blurred due to limited awareness and attribution in both space and cyber domains.

Space cyberattacks are similar to cyber operations against non-space systems. They need access to a system and vulnerability that can be exploited, a malicious payload; and a command-and-control system for communication. Access is gained through the supply chain; the extended land-based infrastructure that sustains space-based assets—including ground stations, terminals, related companies; end-users; and the satellites themselves (Weeden and Samson 2020, p. 9-1). The infiltration of any of it can breach the system. Cyberattacks may be divided into five categories. (1) Attack through faulty or malicious hardware and software components; (2) attack against the links between satellites and ground control stations; (3) attack on terrestrial command and control or data relay stations; (4) attack against user segment of space systems – terminals or devices receiving satellite signal; and closely related, (5) exploitation of satellites links for hacking other targets (Weeden and Samson 2020, pp. 9-1—9-9).

An especially dangerous type of cyberattack is the so-called Advanced Persistent Threats (APTs). APTs *“represent a collection of processes, tools and resources used by certain groups in order to infiltrate specific networks covertly, remain stealthy in the systems over a long period, and exfiltrate data or perform other destructive actions”* (ENISA 2018, p. 87). In the case of space systems, an attacker could gain access to a command-and-control system, enabling him to, for example, *“shut down all communications and permanently damage the satellite by expending its propellant supply or damaging its electronics and sensors”* (Harrison et al, 2020, p. 5).

Hence, cyberattacks can be incorporated into warfare strategies. In 2007, Estonia was struck by a series of cyberattacks which in some cases lasted for weeks and demonstrated strategic importance of focused harmful cyber activity. Though the indirect evidence pointed towards Russia as the offender, there was no clear proof of guilt. It was suggested that the Kremlin organized the attack and vicious gangs joined to support its activity (McGuinness 2017). Only a year later, in 2008, during the Russo-Georgian war Russia launched a cyber operation against Georgia which also widely exploited patriotic Russian “hacktivists” or organized crime, and again, hard evidence for attribution to the Russian government was missing. Though the overall Russian cyber campaign impact was limited, it was the first time the wide-scale offensive cyber operations were mounted in conjunction with conventional military operations (Shakarian 2011). Finally, Russia likely used cyber operations in Ukraine to reach a kinetic effect by targeting Ukrainian critical infrastructure. In December 2015, pro-Russian actors attacked the Ukrainian power grid, temporarily affecting 220,000 residents. The attack is deemed highly sophisticated but intentionally limited since it could allegedly cause permanent damage to power stations (Connell and Vogler 2016, pp. 14-15). The outer space then represents a genuine milestone for extension of cyber operations and the targeting of critical infrastructure.

#### **4. Space cyberattacks and resilience of space systems**

It is fair to say that the record of declassified and public cyberattacks on space systems is limited. However, that is not decreasing its significance; on the contrary, it suggests that vulnerabilities for major space cyberattack and its consequences represent “tabula rasa” and unexplored areas of cyberspace. However, countries such as the United States, Russia, China, North Korea or Iran all revealed a willingness to conduct offensive cyberattacks against non-space systems. Moreover, non-state actors are lurking to exploit commercial space systems' vulnerabilities, which seem to be less resilient than their non-space counterparts (Weeden and Samson 2020, p. 9-2).



Nevertheless, some examples of space cyberattacks are well-known. In 1998, hackers took over U.S.-German ROSAT satellite and aimed solar panels at the sun, frying its batteries and destroying the satellite. In 1999, the U.K.'s SkyNet satellites were hacked for ransom. In 2008, allegedly Chinese attackers gained full control of two NASA satellites for 2 and 9 minutes (Akoto 2020a). In 2007, the Tamil Tiger separatist group managed to disrupt satellite command and control and controlled U.S. commercial satellite's broadcasting (Weeden and Samson 2020, p. 9-3). Moreover, satellites can be corrupted by hardware and software. Faulty components were found; for instance, in Chinese electronics or Russian software packages (Weeden and Samson 2020, p. 9-2). In addition, satellites can be used to gain information from ground systems. Between the years 2008 and 2016, the Russian-led Turla group attacked satellites to get sensitive and confidential data from Western embassies, governments and military institutions. The cyberattack affected 42 countries, including the United States, Germany, and France (Robinson et al, 2018, p. 4). The list of attacks could continue; however, these attacks demonstrate the crucial capabilities and threats of space cyberattacks.

Cyberattack is an inexpensive method that may be utilized by any actor. In 2009, Iraqi insurgents intercepted the U.S. satellite and small scout drone communication by using \$29 SkyGrabber software because the U.S. military did use sufficient security measures (Bardin 2013, p. 1175). In 2020 The U.S. Department of Defense inspector general's office reported that since 2012, cyber vulnerabilities were treated inconsistently without adequate cyber threats mitigation (Pomerleau 2020). A Chatham House research paper on "Cybersecurity of NATO's Space-based Strategic Assets" argues that *"the increasing vulnerability of space-based assets, ground stations, associated command and control systems, and the personnel who manage the systems, has not yet received the attention it deserves."* (Unal 2019, p. 4) This is especially true along in hand with the commercialization of the space sector. During the Iraqi war between the years 2003-2011, there was a 560% increase in the US reliance on commercial satellites for military purposes (Unal 2019, p. 4). Even though the military is generally more cautious about cyber vulnerabilities (Moon 2017, p. 7), their cyber protection thus still registers significant flaws especially in connection to the commercial sphere and progress in fixing vulnerabilities is slow. In 2008, a group of hackers illustrated how easy it is to eavesdrop Iridium company satellites, demonstrating that company, which is, actually, a client to Pentagon, lacks any cyber protection (Porup 2015).

New Space start-up companies and academia with increasing utilization of thousands of small satellites can become an easy target of cyberattack since they tend to underestimate cyber vulnerabilities and such systems are not cyber-hardened (Manulis et al, 2020). As Bardin (2013, p. 1173) concluded, *"[a]s with any growing commercial opportunity, security is less than the primary concern. Economics drives the opportunity."* Though the cyberattacks on space industry were formerly aimed at the ground segment or electronic warfare for espionage and political activity, the New Space era with thousands of new systems located directly in outer space represents a new challenge that can be abused for malicious activity with various motivations and impact (Manulis et al, 2020). Moreover, it is expected that the risks of potential satellite-to-satellite cyberattacks will be on the rise (Falco 2020, p. 7) and orbital systems can thus be cyber-weaponized. Overall, cyberattacks in outer space are not appropriately addressed despite prevailing hazards and more attention should be given to cyber threats (Holmes 2019). For instance, although the U.S. Department of Defense and National Security Agency put some efforts into understanding space cyber threats (Tucker 2019), there is little progress. There are also no cybersecurity standards for satellites, and cyber protection is thus the responsibility of individual companies that operate them (Akoto 2020b).

## **5. Space cyber military doctrine and protection of space assets**

As was indicated, space warfare can be waged to deny access to outer space to the adversary and cripple enemy space capabilities that support terrestrial forces, and also has considerable impact on everyday life. Considering the advantages of cyber weapons- namely affordable cost, limited attribution, debris-limiting properties and a broad scope of impact – cyber warfare strategy could be integrated into space warfare strategies with potentially devastating consequences. A cyberattack could exceed disadvantages of conventional kinetic weapons and *"have the potential to wreak havoc on strategic weapons systems and undermine deterrence by creating uncertainty and confusion"* (Unal 2019, p. 4), which is the key to successful space warfare strategy. The cyber and space domains are increasingly important and are inherently connected; space systems require regular maintenance through vulnerable cyberspace which can be compromised by a hostile entity (Moon 2017, p. 7).

However, the real challenge represents a combination of cyberattack with other means to space systems. In a case of conflict, the kinetic ASAT strike could be accompanied by cyber warfare against both space and ground

systems to increase its magnitude. Moreover, electronic warfare can jam or spoof downlink and uplink radio frequency signals and directed energy such as lasers or high-powered microwaves could dazzle, blind, or directly damage and destroy space systems. Hence, if properly planned, the attacker could gain the upper hand in the conflict already during initial phases and paralyze, or outright destroy its adversary.

The final puzzle is whether cyber threats can be avoided and mitigated. Worth mentioning, Tallin Manual 2.0 is an expert publication on how existing international law applies to cyber operations (NATO Cooperative Cyber Defence Centre of Excellence 2020). Nevertheless, no law will provide insurance against cyberattack.

The White House Space Policy Directive 5 proposed several measures that should be implemented by every satellite owner and operator that deserve to be provided in a full statement:

- 1. “Protection against unauthorized access to critical space vehicle functions. This should include safeguarding command, control, and telemetry links using effective and validated authentication or encryption measures designed to remain secure against existing and anticipated threats during the entire mission lifetime;
- 2. Physical protection measures designed to reduce the vulnerabilities of a space vehicle’s command, control, and telemetry receiver systems;
- 3. Protection against communications jamming and spoofing, such as signal strength monitoring programs, secured transmitters and receivers, authentication, or effective, validated, and tested encryption measures designed to provide security against existing and anticipated threats during the entire mission lifetime;
- 4. Protection of ground systems, operational technology, and information processing systems through the adoption of deliberate cybersecurity best practices. This adoption should include practices aligned with the National Institute of Standards and Technology’s Cybersecurity Framework to reduce the risk of malware infection and malicious access to systems, including from insider threats. Such practices include logical or physical segregation; regular patching; physical security; restrictions on the utilization of portable media; the use of antivirus software; and promoting staff awareness and training inclusive of insider threat mitigation precautions;
- 5. Adoption of appropriate cybersecurity hygiene practices, physical security for automated information systems, and intrusion detection methodologies for system elements such as information systems, antennas, terminals, receivers, routers, associated local and wide area networks, and power supplies; and
- 6. Management of supply chain risks that affect cybersecurity of space systems through tracking manufactured products; requiring sourcing from trusted suppliers; identifying counterfeit, fraudulent, and malicious equipment; and assessing other available risk mitigation measures.” (The White House 2020)

The Space Policy Directive 5 provides an important starting point for developing cybersecurity standards; yet, it is arguably not complete. Sanders and Kordella (2020) from the MITRE Corporation proposed the development of “best practices” in cooperation with government, industry, and other stakeholders to find workable solutions. However, it should be noted that this salutary task would be challenging in terms of both negotiations and implementation without any specialized all-embracing platform and consensus of the private and public sphere.

For this reason, this paper argues that cybersecurity in outer space should be approached in a broad framework of defensive mechanisms. Due to the increasing number of cyber threats, space systems should not focus on mere cybersecurity – protection and avoidance against cyberattack, but on enhanced cyber resilience, ability to operate and mitigate negative consequences of attack (Accenture 2018). It would be somewhat naïve to believe that a severe attack will not happen. Cyber resilience of space assets could be crucial in a case of space conflict, where affected systems could still provide essential functions to mitigate the damage in space and, more importantly, to support terrestrial forces and reduce life loss. Such an ability could also be fostered by increased redundancy of space systems that could be encouraged by building strong space partnerships and cooperation. Strong allies could be deterrent to adversary actors and at the same time provide back-up functions of lost systems. Finally, space actors should build flexible hybrid space architectures to reduce systems vulnerabilities (Erwin 2019).

## **6. Conclusion**

Space and cyber are inherently linked domains the disruption of which can endanger military and civilian terrestrial activities. Cyberspace can be exploited to infiltrate the space domain and seize control over an

adversary's systems. Moreover, cyber threats compared to other counterspace capabilities provide significant advantages such as a low price, limited attribution, flexibility and accessibility. Space systems are vulnerable to cyberattacks, and space cybersecurity is not adequately addressed. Cyberattacks can be lethally integrated into space warfare strategies and wreak havoc upon adversaries by immobilizing and paralyzing their terrestrial forces. The New Space era with the commercialization of the space sector and an increasing number of space systems which lack appropriate cyber defences creates new opportunities for malicious actors. Hence, satellite operators must pay closer attention to cybersecurity issues and enhance their cyber resilience and foster their space architectures.

## Acknowledgements

This study was supported by the Charles University Research Programme "Progres" Q18 - Social Sciences: From Multidisciplinarity to Interdisciplinarity.

## References

- Accenture, 2018. *The Nature Of Effective Defense: Shifting From Cybersecurity To Cyber Resilience*. [ebook] Available at: <[https://www.accenture.com/\\_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Local/en/Accenture-Shifting-from-Cybersecurity-to-Cyber-Resilience-POV.pdf](https://www.accenture.com/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Local/en/Accenture-Shifting-from-Cybersecurity-to-Cyber-Resilience-POV.pdf)> [Accessed 24 December 2020].
- Akoto, W., 2020a. *Hackers Could Shut Down Satellites -- Or Turn Them Into Weapons*. [online] GCN. Available at: <<https://gcn.com/Articles/2020/02/12/hackers-satellites.aspx?Page=1>> [Accessed 23 December 2020].
- Akoto, W., 2020b. *Hackers Could Shut Down Satellites -- Or Turn Them Into Weapons*. [online] GCN. Available at: <<https://gcn.com/Articles/2020/02/12/hackers-satellites.aspx?Page=2>> [Accessed 23 December 2020].
- Analytical Graphics, Inc., 2020. *107,000 Planned Satellites By 2029*. [video] Available at: <<https://www.youtube.com/watch?v=oWB7ZySDHg8>> [Accessed 17 December 2020].
- Bardin, J., 2013. Satellite Cyber Attack Search and Destroy. *Computer and Information Security Handbook*, pp.1173-1181.
- Bowen, B., 2020. *War In Space*. 1st ed. Edinburgh: Edinburgh University Press.
- Connell, M. and Vogler, S., 2016. *Russia's Approach To Cyber Warfare*. [ebook] Center for Naval Analyses. Available at: <<https://apps.dtic.mil/sti/pdfs/AD1019062.pdf>> [Accessed 17 December 2020].
- Dobos, B. and Pražák, J., 2019. To Clear or to Eliminate? Active Debris Removal Systems as Anti-satellite Weapons. *Space Policy*, 47, pp.217-223.
- Dolman, E. (1999). Geostrategy in the space age: An astropolitical analysis. *Journal of Strategic Studies*, 22(2-3), pp.83-106.
- ENISA, 2018. *ENISA Threat Landscape Report 2017*. [ebook] European Union Agency For Network and Information Security (ENISA). Available at: <<https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2017>> [Accessed 17 December 2020].
- Erwin, S., 2019. *Raymond: U.S. Space Command Needs Satellites To Be Built Fast, To Be Survivable - Spacenews*. [online] SpaceNews. Available at: <<https://spacenews.com/raymond-u-s-space-command-needs-satellites-to-be-built-fast-to-be-survivable/>> [Accessed 26 December 2020].
- Falco, G., 2020. When Satellites Attack: Satellite-to-Satellite Cyber Attack, Defense and Resilience. In: *ASCEND 2020 Virtual Conference*. [online] Available at: <<https://doi.org/10.2514/6.2020-4014>> [Accessed 23 December 2020].
- Georgescu, A., Gheorghe, A., Piso, M. and Katina, P., 2019. *Critical Space Infrastructures: Risk, Resilience And Complexity*. Springer.
- Harrison, T., Johnson, K., Roberts, T., Way, T., Young, M. and Faga, M., 2020. *Space Threat Assessment 2020*. [ebook] Washington, DC: Center for Strategic and International Studies. Available at: [https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/200330\\_SpaceThreatAssessment20\\_WEB\\_FINAL1.pdf?6sNra8FsZ1LbdVj3xY867tUVu0RNHw9V](https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/200330_SpaceThreatAssessment20_WEB_FINAL1.pdf?6sNra8FsZ1LbdVj3xY867tUVu0RNHw9V) [Accessed 16 December 2020].
- Havercroft, J. and Duvall, R. (2009). Critical astropolitics: The geopolitics of space control and the transformation of state sovereignty. In: N. Bormann and M. Sheehan, ed., *Securing Outer Space*. New York: Routledge, pp.42-58.
- Holmes, M., 2019. *The Growing Risk Of A Major Satellite Cyber Attack*. [online] Via Satellite. Available at: <<http://interactive.satellitetoday.com/the-growing-risk-of-a-major-satellite-cyber-attack/>> [Accessed 23 December 2020].
- Johnson-Freese, J. (2007). *Space as a strategic asset*. New York: Columbia Univ. Press.
- Kleinberg, H. (2007). On War in Space. *Astropolitics*, 5(1), pp.1-27.
- Lourenço, M. and Marinou, L., 2020. *Main Incidents In The EU And Worldwide*. [ebook] European Union Agency for Cybersecurity (ENISA). Available at: <<https://www.enisa.europa.eu/publications/enisa-threat-landscape-2020-main-incidents>> [Accessed 17 December 2020].
- Manulis, M., Bridges, C., Harrison, R., Sekar, V. and Davis, A., 2020. Cyber security in New Space. *International Journal of Information Security*, [online] Available at: <<https://link.springer.com/article/10.1007/s10207-020-00503-w>> [Accessed 23 December 2020].
- McGuinness, D., 2017. *How A Cyber Attack Transformed Estonia*. [online] BBC News. Available at: <<https://www.bbc.com/news/39655415>> [Accessed 17 December 2020].

- Moon, M., 2017. *THE SPACE DOMAIN AND ALLIED DEFENCE*. [ebook] NATO Parliamentary Assembly. Available at: <<https://www.nato-pa.int/download-file?filename=sites/default/files/2017-11/2017%20-%20162%20DSCFC%2017%20E%20rev%201%20fin%20-%20SPACE%20-%20MOON%20REPORT.pdf>> [Accessed 25 December 2020].
- NATO Cooperative Cyber Defence Centre of Excellence. 2020. *Tallinn Manual 2.0*. [online] Available at: <<https://ccdcoe.org/research/tallinn-manual/>> [Accessed 25 December 2020].
- NATO. 2019. *NATO Will Defend Itself*. [online] Available at: <[https://www.nato.int/cps/en/natohq/news\\_168435.htm?selectedLocale=en](https://www.nato.int/cps/en/natohq/news_168435.htm?selectedLocale=en)> [Accessed 16 December 2020].
- Pomerleau, M., 2020. *The Pentagon Is Handling Cyber Vulnerabilities Inconsistently*. [online] Fifth Domain. Available at: <<https://www.fifthdomain.com/dod/2020/03/17/the-pentagon-is-handling-cyber-vulnerabilities-inconsistently/>> [Accessed 24 December 2020].
- Porup, J., 2015. *It's Surprisingly Simple To Hack A Satellite*. [online] Vice. Available at: <<https://www.vice.com/en/article/bmjg5a/its-surprisingly-simple-to-hack-a-satellite>> [Accessed 24 December 2020].
- Robinson, J., Šmuclerová, M., Degl'Innocenti, L., Perrichon, L. and Pražák, J. (2018). *EUROPE'S PREPAREDNESS TO RESPOND TO SPACE HYBRID OPERATIONS*. [ebook] PSSI. Available at: [https://www.pssi.cz/download//docs/8252\\_597-europe-s-preparedness-to-respond-to-space-hybrid-operations.pdf](https://www.pssi.cz/download//docs/8252_597-europe-s-preparedness-to-respond-to-space-hybrid-operations.pdf) [Accessed 16 December 2020].
- Rousseau, D. and Garcia-Retamero, R., 2007. Identity, Power, and Threat Perception. *Journal of Conflict Resolution*, 51(5), pp.744-771.
- Shakarian, P., 2011. *The 2008 Russian Cyber-Campaign Against Georgia*. [ebook] Military Review. Available at: <[https://www.armyupress.army.mil/Portals/7/military-review/Archives/English/MilitaryReview\\_20111231\\_art013.pdf](https://www.armyupress.army.mil/Portals/7/military-review/Archives/English/MilitaryReview_20111231_art013.pdf)> [Accessed 17 December 2020].
- Sun Tzu, 2021. *The Art Of War*. [ebook] The Internet Classics Archive. Available at: <<http://classics.mit.edu/Tzu/artwar.html>> [Accessed 5 January 2021].
- The White House. 2020. *Memorandum On Space Policy Directive-5—Cybersecurity Principles For Space Systems | The White House*. [online] Available at: <<https://www.whitehouse.gov/presidential-actions/memorandum-space-policy-directive-5-cybersecurity-principles-space-systems/>> [Accessed 25 December 2020].
- Tucker, P., 2019. *The NSA Is Studying Satellite Hacking*. [online] Defense One. Available at: <<https://www.defenseone.com/technology/2019/09/nsa-studying-satellite-hacking/160009/>> [Accessed 23 December 2020].
- Unal, B., 2019. *Cybersecurity Of NATO'S Space-Based Strategic Assets*. [ebook] Chatham House. Available at: <<https://www.chathamhouse.org/sites/default/files/2019-06-27-Space-Cybersecurity-2.pdf>> [Accessed 24 December 2020].
- Union of Concerned Scientists. 2004. *International Legal Agreements Relevant To Space Weapons*. [online] Available at: <<https://www.ucsusa.org/resources/legal-agreements-space-weapons>> [Accessed 17 December 2020].
- United Nations Office for Disarmament Affairs. 2020. *Treaty Banning Nuclear Weapon Tests In The Atmosphere, In Outer Space And Under Water*. [online] Disarmament.un.org. Available at: <[http://disarmament.un.org/treaties/t/test\\_ban/text](http://disarmament.un.org/treaties/t/test_ban/text)> [Accessed 17 December 2020].
- Visner, S. and Kordella, S., 2020. *Cyber Best Practices For Small Satellites*. [online] The MITRE Corporation. Available at: <<https://www.mitre.org/publications/technical-papers/cyber-best-practices-for-small-satellites>> [Accessed 25 December 2020].
- Weeden, B. and Samson, V., 2020. *Global Counterspace Threats: An Open Source Assessment*. [ebook] Secure World Foundation. Available at: [https://swfound.org/media/206970/swf\\_counterspace2020\\_electronic\\_final.pdf](https://swfound.org/media/206970/swf_counterspace2020_electronic_final.pdf) [Accessed 16 December 2020].

# e-Health as a Target in Cyberwar: Expecting the Worst

Samuel Wairimu

Department of Mathematics and Computer Science, Karlstad University, Universitetsgatan 2, Sweden

[samuel.wairimu@kau.se](mailto:samuel.wairimu@kau.se)

DOI: 10.34190/EWS.21.054

**Abstract:** Healthcare organisations have become a key target for attackers as evidenced by the global increase in cyber-attacks. These cyber-attacks are attributed to various attackers who differ in motivations and skills, with the common motivation being financial gain due to the rich personal data contained in patients' health records. But what would happen if the motivation changed? What would happen if the motivation is driven by targeting key people, mass exploitation or taking lives? What would happen if a strategic cyber-attack knocks out a society's critical infrastructure? This article investigates the possibility of targeting e-Health in the context of cyberwar. It assesses the privacy in healthcare and compares the consequences and impact of conventional cyber-attacks within the healthcare sector, against the consequences and impact of cyberwar on the same. The outcome indicates that e-Health in the context cyberwar could result to active reconnaissance of patient records, which could lead to the targeting of key and influential people through Personally Identifiable Information (PII), mass exploitation, and personal attacks derived from Personal Health Information (PHI), which could result to irreversible damage or death.

**Keywords:** e-Health, privacy, cyberwar, cyber-attack, critical-infrastructure, healthcare

## 1. Introduction

The delivery of healthcare services has been highly reshaped through the advances in e-Health. This has created vast and positive potential in terms of promoting service delivery through the redefinition of patient and healthcare professional relationship, advancing clinical outcomes (e.g., through the improvement of clinical-decision making) and providing both quality and cost-effective health care services. As such, governments across the globe are contributing on the implementation of e-Health within their respective healthcare systems due to the aforementioned reasons, among others (Ross *et al.*, 2016). For example, Sweden laid out a vision for e-Health, where the country aims to be the best in the globe by year 2025 in terms of leveraging e-Health to provide improved and equal health care services while promoting independence among healthcare consumers (Wickström, Regner and Micko, 2017).

However, while e-Health offers such advantages, cyber-attacks against the healthcare organisations are increasing exponentially (Argaw *et al.*, 2019). This can be noted from the U.S. Department of Health and Human Services Breach Portal where data breaches within healthcare organisations are published (HHS, 2020). The EU has not been spared either as reports indicate that healthcare organisations have been targeted during the COVID-19 pandemic (Lepassaar, 2020). For example, in the case of Brno University Hospital in the Czech Republic where an attack led to the unavailability of critical services and re-routing of emergency cases (Cimpanu, 2020).

**Table 1:** Overview of attackers and their targets in the context of cyber-attacks in e-Health. These attacks can be directed toward a specific victim or could be random based on the attacker's motivations. Image adopted from ISE (2016)

Adversary	Patient Health		Patient Records	
	Targeted (Specific Victims)	Untargeted (Indiscriminate)	Targeted (Specific Victims)	Untargeted (Indiscriminate)
Individual/Small Group				Yes
Political Groups/Hacktivists			Yes	
Organized Crime	Yes		Yes	Yes
Terrorism/Terrorist Organisation	Yes	Yes		
Nation States	Yes	Yes	Yes	Yes

Conventionally, these attacks emanate from different attackers. Table 1 shows an overview of identified attackers that target healthcare organisations with the aim of compromising either patients records or health,

or both. These attackers have contrasting motives and skills as highlighted in Table 2. While all these are relevant, the focus is turned to Nation State attackers.

**Table 2:** Types of attackers with their skills and motivations

Adversary	Skills and Motivation
Individual/Small Group attackers	Driven by financial gain, which is a key incentive when it comes to healthcare cyber-attacks (Martin <i>et al.</i> , 2017). This is because patient records contain PII, such as Credit Card Information and Social Security Numbers (SSN) and PHI - which is deemed more valuable as it can be used for medical identity theft and insurance fraud (Coventry and Branley, 2018). Hence, these attackers are indiscriminate and mostly depend on unsophisticated means when launching an attack (ISE, 2016)
Political Groups/Hacktivists	In their paper, Coventry and Branley (2018) state that patient records contain information that can be leveraged in politics. Therefore, these groups can use such information to embarrass or coerce a well-known individual in the political arena for their own personal gain. In addition, they tend to less skilled and hence employ skilled people.
Organized Crime	Relies on highly experienced attackers and can be indiscriminate or target specific victims by influencing the patient health or records as indicated in Table 1. Hence, they are normally driven by financial benefits or other illegal activities such as extortion (ISE, 2016).
Terrorism/Terrorist organisations	Driven by causing fear or harm to either a specific victim or random individuals by attempting to compromise their health. Research conducted by the Independent Security Evaluators (ISE) (ISE, 2016) indicate that such attackers do not demonstrate high skills as those from organized crime or nation state attackers.

### 1.1 Nation State attackers

When it comes to Nation State (henceforth called state-sponsored) attackers, their motivations change and the level of sophistication demonstrated differs from the aforementioned attackers. In addition to the vast patient records, the vulnerabilities inherent within e-Health makes the healthcare a lucrative sector for attackers (Martin *et al.*, 2017), for example, software vulnerabilities. In fact, McGraw (2013) argues that such ICT integrated systems not only harbour a lot of vulnerabilities that increase cyber threats, but our growing dependence on them is a key factor that makes cyberwar against such systems unavoidable. Indeed, it can be argued that with such vulnerabilities, which translates to a poor healthcare security posture, the issue of privacy arises due to the potential compromise of patient records. Hence, in the current threat landscape where state-sponsored attackers are hitting healthcare organisations with cyber-attacks, the consequences and the impact can be vast and devastating. For example, at a strategic level, a cyber-attack against e-Health could compromise the integrity of data systems, privacy of patients (through the exposure of sensitive personal and health data - which might include key personnel within the government or military), mass disruption of critical facilities, or mass exploitation.

From Table 1, it can be noted that state-sponsored attackers are indiscriminate, with the probability of conducting mass exploitation. However, like Political Groups/Hacktivists or Organized Crime, state-sponsored attackers tend to go for targeted victims, and hence, choose specific healthcare organisations. Nevertheless, while the Political Groups/Hacktivists would target even small healthcare organisations, state-sponsored attackers would not go for this (ISE, 2016). To bring about a large impact, they could take out not only one, but many big healthcare organisations within a target state. Furthermore, they can go for patients records with the aim of targeting a patient’s health through the modification of medical records or prescriptions (Gross, Canetti and Waismel-Manor, 2015).

In his paper, Jan Kallberg explains how a strategic cyber-attack against a targeted society's critical infrastructure can knock out one or all of the five pillars identified by Dwight Waldo, that is, legitimacy, authority, institutional knowledge, bureaucratic control and confidence (Kallberg, 2016). Hence, this could not only develop a belief that the government is incapable of running the country, but the citizens of that particular country can lose confidence towards the healthcare organisations and the government.

From this, one fundamental question arises: *What would be the impact of weaponizing e-Health in the context of cyberwar?* To explore this, the article sets this complex problem by examining the concept of e-Health as a critical infrastructure within the society, and its potential use and abuse in the context of cyberwar in relation to privacy.

## **2. Background**

Undoubtedly, e-Health is playing a major role in today's society. As mentioned earlier, there have been a number of potential benefits that have been realised through the introduction of e-Health within the healthcare sector.

Traditionally, the delivery of healthcare services depended on offline files and paper-based records, which led to many discrepancies and inefficiencies (Gaddi, Capello and Manca, 2013). This was later changed and improved through the implementation of e-Health. According to Ossebaard and Van Gemert-Pijnen (2016), e-Health refers to the use of information and communication technologies to support health, well-being and the healthcare system. These technologies span from those offering storage of Electronic Health Records (EHRs), medical devices (those that keep track of health and administer medication - in addition to wearables and implantable health devices), mHealth applications, and telemedicine (Coventry and Branley, 2018), to the reinforcement of secondary processes of care, such as, booking of appointments and case management (Ossebaard and Van Gemert-Pijnen, 2016).

As a result, e-Health supports consumers with improved access and control over their health by emphasising self-management, involvement, and transparency (Erlingsdóttir and Sandberg, 2016), while at the same time providing healthcare professionals with an opportunity to make clinical decisions and to improve consumers' welfare (Gaddi, Capello and Manca, 2013; Kreps and Neuhauser, 2010). As such, e-Health - which is deemed as a critical infrastructure - can destabilise the population and at the same time put the national security at risk when hit at a strategic level, as it is often, among other critical infrastructures (e.g., Transportation, Water systems and Energy), considered a preferred military target (Walker-Roberts, Hammoudeh and Dehghantaha, 2018).

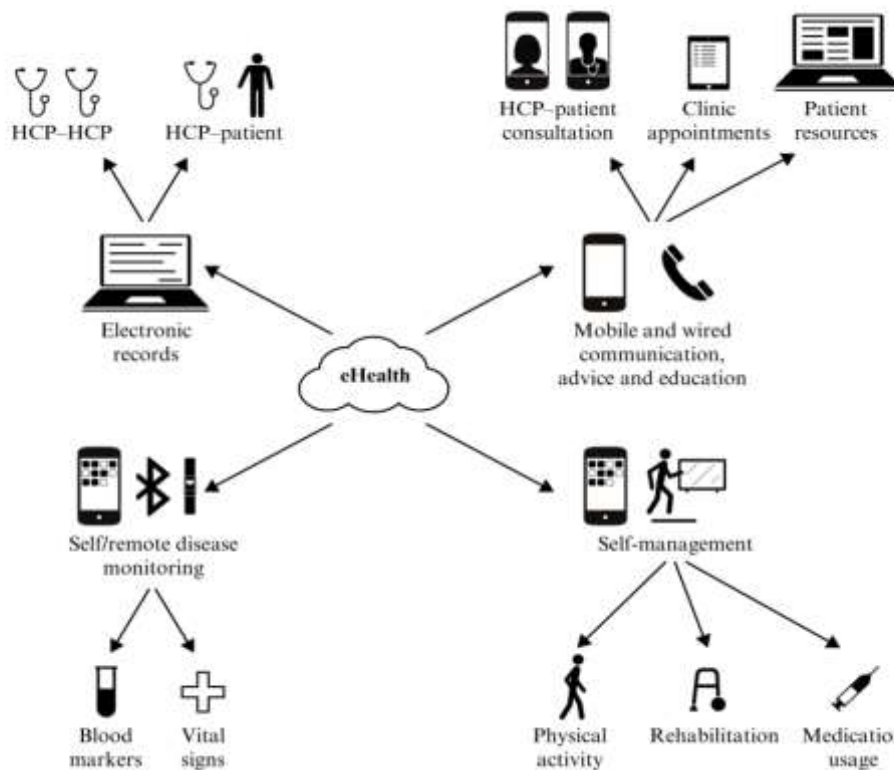
## **3. Privacy in healthcare**

In the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR), the privacy of patients, especially in regards to their health data, is deemed as a core principle. Providing sensitive information to a healthcare professional is regarded as necessary in order to give proper diagnosis and prevent unnecessary drug interactions among other things (Appari and Johnson, 2010), and protecting this information is critical as its leakage would lead to either discrimination or social stigma towards the patient. Armed with patients' records, an attacker could use it to blackmail an individual, thus causing psychological or physiological effects or intimidate individual with harm (Gross, Canetti and Waismel-Manor, 2015). As such, privacy is considered as a critical element in the healthcare. However, with the advances in e-Health, the notion of privacy in healthcare arises due to the inherent vulnerabilities in its cybersecurity infrastructure. Such vulnerabilities have led to massive data breaches, which expose the PHI of patients thus affecting their privacy.

While it is acknowledged that there has been major improvement in the healthcare organisations attributed to e-Health, Coventry and Branley (2018) argue that its security posture is wanting; in fact, it can be argued that the privacy of a patient could be based on the security measures implemented, which more than ever seems to be not the case due to the rising number of cyber-attacks against the healthcare. This has been attributed to an increase in inter-connectivity, which exposes the sector to not only common, but also new to new vulnerabilities that weaken the security posture of the healthcare system thus risking health data breach and manipulation of health devices (Williams and Woodward, 2015). Additionally, taken in isolation, disruptive technologies introduced in e-Health tend to introduce unanticipated vulnerabilities that create more cyber threats. According to Liff (2012), attacks that use these vulnerabilities in computer network could be exploited in cyberwar. This would create panic in the society (either through the exposure of patient records, and disruption of critical health infrastructure), loss of confidence with the authorities and damage of reputation to the organisation or health professional.

#### 4. Attack vectors

As identified in the background section, e-Health includes a number of technologies that span from those being used by stakeholders, that is patients and healthcare professionals within the healthcare sector, as identified in Figure 1.



**Figure 1:** Overview of e-Health Technologies used within the healthcare organisations and with patients. Image adapted from Marziniak *et al.*, (2018)

These technologies collect PHI from patients, through sensors, that is later transmitted to a wireless device, for instance mobile phone (Zhang *et al.*, 2016). From there, the patient sends the data to a healthcare professional where the data is processed and integrated into Electronic Health Records (EHRs). The data is later accessed and shared between healthcare professionals depending on the purposes. This integration of ICT and massive interconnectivity has opened doors to vulnerabilities (Coventry and Branley, 2018) that threaten the privacy of patients by giving rise to a wider attack surface. Some of these vulnerabilities that are specific in this case are: authentication challenges which would allow the exploitation of Bluetooth enabled devices (Zubair *et al.*, 2019), zero-day vulnerabilities, insecure communication protocols (Zhang *et al.*, 2016) and unpatched vulnerabilities.

In addition to this, the sector has intersected with a number of organisations (which also play the role of stakeholders within the healthcare sector) (Walker-Roberts, Hammoudeh and Dehghantaha, 2018), for example, government, health insurances and pharmacies, which further enlarges the attack surface that state sponsored attackers could take advantage of. For example, attackers could take advantage of a less secure pharmacy through with the aim of targeting a larger healthcare organisation through a supply chain attack.

With an enlargement of the attack surface, comes the increase of attack vectors, which need to be addressed to prevent massive disclosure of information for the sake of social disruption or surveillance.

The following are some of the attack vectors identified together with their descriptions in the context of e-Health as a target in cyberwar:

- 1. Phishing - The use of phishing could be used when an attacker aims to gain access to sensitive patient information. For example, the phishing attack that occurred in Jefferson Healthcare (Jefferson Healthcare, 2021) thus affecting 2,550 patients. While this attack might seem trivial to a cyberwarrior, a cyberwarrior can use this vector to access the information of specific individuals.



- 2. Distributed Denial of Service (DDoS) – This has been identified by Abbas et al (2016) as a powerful attack on both the availability of PHI and the services offered by the healthcare professionals. Such an attack, while it might not lead to any privacy issue, could lead to denial of critical health information. In a cyberwarrior context, this can be used to destabilise the society or overwhelm the healthcare system.
- 3. Unpatched vulnerabilities – Legacy systems tend to have unpatched vulnerabilities as they are no longer supported. In the case of the NHS attack, the use of legacy systems and the availability of unpatched vulnerabilities led to the spread of the WannaCry ransomware. While it was not the target as highlighted in the investigative report (Morse, 2018), this indicates how healthcare organisations can be victims of state sponsored attackers.
- 4. Eavesdropping – Transmitting unencrypted PHI over the internet could lead to eavesdropping; as such, an attacker can be able to tell what kind of ailment the patient is suffering from based on the intercepted packets (Zubaydi et al., 2015). Hence, an attacker can modify the transmissions, which could lead to improper treatment of the patient, or disclose sensitive patient data.

In their paper, Limba *et al.*, (2019) identify and place cyber-attacks against critical infrastructures in five major threat groups. These are: Data manipulation, service unavailability, leakage of personal information, unauthorised access of resources and physical destruction. By using the listed attack vectors, four of these cyber-attacks against e-Health could put the privacy of user at risk or possibly lead to irreversible damage. These are:

- 1. Data manipulation - Insecure communications channels could be intercepted thus providing an avenue for either modification or eavesdropping of patient's health data.
- 2. Leakage of personal information - An attacker could gain access to PII, including health data, which they could later disclose.
- 3. Unauthorised access of resources - Attackers could gain access and steal medical resources, which could include research data e.g., an intelligence group trying to steal COVID-19 research data from Canada, the British and the US (Palca and Myre, 2020).
- 4. Service unavailability - While this might not have a direct impact on the privacy of patients, unavailability of services in a healthcare organisation can lead to irreversible consequences, for instance death. An example of this is the cyber-attack at Düsseldorf University Hospital (Germany) that led to the death of a patient as a result of disruption of critical services that were needed for her prompt treatment (Tidy, 2020).

## **5. Impact of cyberwar against e-Health in relation to personal data**

Sevis and Seker (2016) state that attackers compromise critical systems with the intention of pilfering, damaging or taking control of critical information. This is supported by Limba *et al.*, (2019) as identified in the aforementioned five major threat groups. In the context of e-Health, patient records contain extensive information, and unlike financial data, resetting certain identifiers, for example, name, address and SSN, remains impossible (Martin *et al.*, 2017). This makes it a potential target for attackers at different levels as identified in Table 1. Additionally, it contains a full medical history of an individual, which indicates the type of medication they are on, and if they have a medical device to support their health or not. This information can be used to target a patient's health thus causing harm.

To highlight the impact of cyberwar against e-Health, a comparison is made between a non-state sponsored (e.g., Individual/Small Group) and state-sponsored attacker. Hence, based on the identified attack vectors, the exploitation of e-Health is conducted for several motives and with different outcomes. This could range from disgruntled employee, accidental data breach, to low-level malicious data breaches. With the health data containing an extensive source of valuable information, Martin *et al.*, (2017) identifies that financial gain is the key motive for compromising healthcare systems. As such, a cyber-attack launched by non-state sponsored attackers would have the following consequences and impact listed in Table 2.

However, in an age where global awareness of cyberwarfare has increased abruptly (Arquilla, 2013), cyber-attacks could not only be driven by political or financial gain, but by the ability to take lives (Coventry and Branley, 2018) and advance the interest of the particular nation state that has initiated the attack (Ablon, 2018). As such, weaponization of health data and patients' records can have severe consequences and impact listed in Table 3.

**Table 3:** Consequences and impact of low-level, non-state sponsored attacks against e-Health

Consequences	Impact
Active surveillance – collecting patients records, which contains both PII and PHI.	The disclosure of patient’s PII and PHI could result from random people getting targeted for medical identity theft and insurance fraud (Coventry and Branley, 2018).
Personal attacks derived from PHI.	Unauthorized access to PHI could lead to blackmailing or cyberbullying due to sensitive information documented. These could result to physiological and psychological effects or financial harm.
Disruption of medical services	While this might have an indirect effect on privacy, the disruption of medical services could result in cancellation of appointments. Further, this could cause indirect deaths associated with cyber-attacks against critical services, for example in Düsseldorf’s University Clinic in Germany.

**Table 4:** Consequences and impact of weaponizing PHI and PII in the context of cyberwar against e-Health.

Consequences	Impact
Active Reconnaissance – collecting patients’ records for enemy intelligence	Due to the disclosure of PII from health data, key people can be targeted (either within the government or military), e.g., The Trident Juncture 18 (Hughes, 2018). Also, the disclosed PII can be used to build a database of targets.
Personal attacks derived from PHI	Disclosed e-prescriptions and personal medical records can be altered (Gross, Canetti and Waismel-Manor, 2015) with nefarious intentions of putting the patients health and safety at risk. Disclosure of famous people medical files – for example, the case of medical files disclosed from the World Anti-Doping Agency (BBC, 2016). Data manipulation of vulnerable medical devices, e.g., insulin pumps (Fu and Blum, 2013) that could cause irreversible damage. Finally, disclosed PHI could contain information that could be used for blackmailing purposes, which could lead to psychological effects.
Mass disruption of medical services	While there could be some possibility of data breach, the disruption of medical services could lead to mass interruptions of appointments (as in the case of NHS). Disruption of medical facilities could lead to fatalities, and the disruption of care and emergency services, which could lead to public unrest. This can further cause the population to lose trust with their government and healthcare organisations.
Mass defacement of health websites	Access and modifications of resources with the aim of spreading propaganda, for example in the case of NHS websites (Southall, 2017), could lead to social disruption. Also, the unavailability of information could cause anger, mass confusion and scare.

As indicated in Table 3, the exploitation of e-Health with the resources and skills of cyberwar causes an impact on the physical domain - which are ultimately felt on the civilian, military and government spheres. Following the same approach applied by Fritsch (2020), the consequences of exploiting e-Health in the context of cyberwar, would have privacy issues as shown in Table 4.

**Table 5:** Disclosing sensitive data from the e-Health digital sphere in the context of cyberwar could have severe privacy issues when exploited

e-Health Digital Sphere	Physical Sphere
PII – This contains data that could be used to potentially identify a healthcare consumer.	The impact of disclosing such data would be creating a database of targets by accessing their names, addresses, locations, etc.
PHI – This contains an entire history of a patient’s medical history.	The impact of disclosing such data would result in learning the behaviour of a patient, insurance information, location, medical histories, possible allergies, etc. This could potentially be used for nefarious purposes.
System Security Data – This contains data such as access control data e.g., login credentials	The impact of this would be the disclosure of staff login credentials that would cause further leakage of patients’ data or modification and possible privilege escalation with the intent of taking control over the entire critical infrastructure information.

## 6. Mitigating cyberwar in e-Health

Having established the impact of exploiting patients' records in the context of cyberwar, the situation highlights a dire need to improve the security and privacy of e-Health. Research conducted by ISE (2016) identifies that healthcare organisations devalue the motivation and sophistication of state sponsored attackers and only address the low-level attackers. As such, it is imperative for healthcare organisations to acknowledge state sponsored attackers by understanding their motivation and recognising their profiles. By doing this, health organisation would be able to improve their state to prevent state sponsored attackers.

Hence, certain measures can be taken to mitigate cyberwar in the context of e-Health. For instance, healthcare organisations, for example, the NHS (UK) and HHS (US), are now investing in cyber and information security after the WannaCry. Granted that the healthcare sector implements existing laws (e.g. HIPAA or invest in the current standards (e.g., ISO/IEC 27001:2017) and frameworks (e.g., NIST-CSF or NIST-PF); would these be enough to prevent cyberwar in the context of e-Health?

It can be argued that these laws, standards, and frameworks are rather relevant; however, in the context of cyberwar and information security in healthcare, a number of them would come in handy. For example, ISO/IEC 27032:2012 that describes guidelines for cybersecurity, provide controls for addressing cyber risks, including controls for cyber organised criminals (International Organization for Standardization, 2012). Further, the standard highlights within its scope the guidance required for improving, among others, the cybersecurity of critical information infrastructure protection. As such, this standard can play an important and urgent role when mitigating the threat of state sponsored attackers. In addition, ISO/IEC 27799:2016 could be implemented in this context. This standard describes specific guidelines for information security management in health using ISO/IEC 27002 (Ac, 2008). This is a special type of standard that focuses entirely on the privacy and security of e-Health by ensuring best practices are followed withing the healthcare sector. By doing so, health information data, such as PHI, medical research data and other sensitive data concerned with system security is protected from malicious actors, which include the state sponsored attackers.

However, while these standards are urgently recommended for better privacy and security practices, there needs to be implementation of other cybersecurity measures to support the above. According to McGraw (2013), one way to prevent this is to build security in the system that factors in skilled and resourceful attackers with high levels of intent in consideration. This could be done by applying Article 25 of the GDPR, which ensures data protection by design and by default, and Article 32, which ensures appropriate levels of security when processing personal data. Furthermore, calculating possible risks by cybermapping all hardware and software within a critical infrastructure could protect critical infrastructure like e-Health from cyber threats. Identifying, calculating and treating risks within the healthcare sector could play part in the reduction of privacy threats. During this process, it is imperative to identify an appropriate approach and utilise the right risk metrics. Healthcare organisations can also go for high privacy impact assessments, which can be used to prevent privacy risks, not only in the context of non-state sponsored attackers, but also in the context of state sponsored attackers. In addition, improving the identification and management of highly advanced cyber incidents and attacks against critical infrastructures, for example, in the case of the European collaborative early warning system ECOSSIAN (Kaufmann *et al.*, 2015), is also imperative in preventing the impact of cyberwar in e-Health.

## 7. Discussion

Regardless of the motivation, the impact and consequences of hitting a healthcare organisation with a cyber-attack are dire. However, differences can be noted when each attacker is taken in isolation. For example, low-level skills, non-sponsored attackers tend to compromise e-Health with the intention of financial gain as patients' records hold valuable data. In addition to this, there is normally disruption of medical services.

However, when it comes to highly skilled and sophisticated state sponsored attackers, the consequences and impact become devastating. It can be noted from Table 3 that state sponsored attackers tend to have broad motivations when they compromise e-Health. Furthermore, when they compromise e-Health, the impact is felt on the physical sphere. For example, targeting a person through their PHI can result in learning sensitive information about the individual which could later be used for nefarious purposes. The exploitation of e-Health can further lead to mass exploitation, which would either lead to mass surveillance of the population or possibly interfere with elections.

While some strategic cyber-attacks are directed towards e-Health, some of the attacks experienced are because of cyber-collateral damage (Romanosky and Goldman, 2016). For instance, the WannaCry cyber-attack which paralysed e-Health across the NHS (UK) resulting in cancellation of surgeries, hospital diversion of emergencies and unavailability of patient records in both England and Scotland (Mattei, 2017). Initial investigations indicate that the NHS was not the specific target (Morse, 2018) hence showing that cyber-collateral damage can have adverse effects on e-Health in the context of cyberwar. Further, according to Fritsch and Fischer-Hübner (2018) "*future Battlefield of Things will be the weaponization of civilian or dual-use infrastructure.*" This suggests that through interlinking of these infrastructures, healthcare organisations can be caught in the crossfire thus having a destructive impact on an entire nation (Walker-Roberts, Hammoudeh and Dehghantanha, 2018) as highlighted in Table 3.

To prevent direct and indirect possible state-sponsored attacks against e-Health, which could lead to the weaponization of PHI and PII, it is recommended that healthcare sectors together with the relevant stakeholders, play part in ensuring that the protection of critical information infrastructure. This can be done by urgently applying a number of standards identified in Section 6, for example, ISO/IEC 27032:2012 and ISO/IEC 27799:2016. In addition to this, risk assessments should be conducted, taking in mind the profiles and motivations of sophisticated state sponsored attackers.

## 8. Conclusion

The question of e-Health as a target in cyberwar is one to ponder on. It is widely acknowledged that e-Health carries a lot of patients' records (which include Personal Health Information and Personally Identifiable Information), and when breached, it could affect millions of not thousands. Such information is normally targeted by attackers due to its value in the black market.

However, when the motivation goes beyond financial gain, health data, including PII, could be weaponised as indicated in Table 3. This could lead to devastating consequences that could lead to a number of privacy impacts within the physical domain as indicated in Table 4. Hence, to prevent this, several mitigations highlighted in section 6, for example, ISO/IEC 27799:2016, need to be implemented to avoid the worst from happening.

## References

- Abbas, H., Latif, R., Latif, S. and Masood, A. (2016) 'Performance evaluation of Enhanced Very Fast Decision Tree (EVFDT) mechanism for distributed denial-of-service attack detection in health care systems', *Annals of Telecommunications*, 71(9), pp. 477-487.
- Ablon, L. (2018) 'Data thieves: The motivations of cyber threat actors and their use and monetization of stolen data'.
- Ac, A. (2008) *Health informatics-Information security management in health using ISO/IEC 27002*. ISO.
- Appari, A. and Johnson, M. E. (2010) 'Information security and privacy in healthcare: current state of research', *International journal of Internet and enterprise management*, 6(4), pp. 279-314.
- Argaw, S. T., Bempong, N.-E., Eshaya-Chauvin, B. and Flahault, A. (2019) 'The state of research on cyberattacks against hospitals and available best practice recommendations: a scoping review', *BMC medical informatics and decision making*, 19(1), pp. 1-11.
- Arquilla, J. (2013) 'Twenty years of cyberwar', *Journal of Military Ethics*, 12(1), pp. 80-87.
- Cimpanu, C. (2020) *Czech Hospital hit by cyberattack while in the midst of a COVID-19 outbreak*.
- Coventry, L. and Branley, D. (2018) 'Cybersecurity in healthcare: a narrative review of trends, threats and ways forward', *Maturitas*, 113, pp. 48-52.
- Erlingsdóttir, G. and Sandberg, H. (2016) 'eHealth opportunities and challenges: a white paper'.
- Fritsch, L. (2020) 'Identity Management as a target in cyberwar', *Open Identity Summit 2020*.
- Fritsch, L. and Fischer-Hübner, S. (2018) 'Implications of Privacy & Security Research for the Upcoming Battlefield of Things', *Journal of Information Warfare*, 17(4), pp. 72-87.
- Gaddi, A., Capello, F. and Manca, M. (2013) *eHealth, care and quality of life*. Springer.
- Gross, M. L., Canetti, D. and Waismel-Manor, I. (2015) 'The Psychological & Physiological Effects of Cyberwar', *Binary Bullets: The Ethics of Cyberwarfare*, pp. 157-76.
- HHS (2020) *Breach Portal: Notice of the Secretary of HHS Breach of Unsecured Protected Health Information*. Available at: [https://ocrportal.hhs.gov/ocr/breach/breach\\_report.jsf](https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf) (Accessed: 06th January 2021 2021).
- International Organization for Standardization, I. E. C. (2012) 'ISO/IEC 27032: 2012—Information technology—Security techniques—Guidelines for cybersecurity'.
- ISE 2016. *Securing Hospitals: A Research Study and Blueprint*.
- Jefferson Healthcare (2021) *Jefferson Healthcare contain data breach, notifies those affected*.
- Kallberg, J. (2016) 'Strategic cyberwar theory-A foundation for designing decisive strategic cyber operations', *The Cyber Defense Review*, 1(1), pp. 113-128.

- Kaufmann, H., Hutter, R., Skopik, F. and Mantere, M. (2015) 'A structural design for a pan-European early warning system for critical infrastructures', *e & i Elektrotechnik und Informationstechnik*, 132(2), pp. 117-121.
- Kreps, G. L. and Neuhauser, L. (2010) 'New directions in eHealth communication: opportunities and challenges', *Patient education and counseling*, 78(3), pp. 329-336.
- Lepassaar, J. (2020) *Healthcare Cybersecurity in the Time of COVID-19*. Available at: <https://healthmanagement.org/c/healthmanagement/issuearticle/healthcare-cybersecurity-in-the-time-of-covid-19> (Accessed: 6th January 2020 2021).
- Liff, A. P. (2012) 'Cyberwar: a new 'absolute weapon'? The proliferation of cyberwarfare capabilities and interstate war', *Journal of Strategic Studies*, 35(3), pp. 401-428.
- Limba, T., Plêta, T., Agafonov, K. and Damkus, M. (2019) 'Cyber security management model for critical infrastructure'.
- Martin, G., Martin, P., Hankin, C., Darzi, A. and Kinross, J. (2017) 'Cybersecurity and healthcare: how safe are we?', *Bmj*, 358, pp. j3179.
- Marziniak, M., Brichetto, G., Feys, P., Meyding-Lamadé, U., Vernon, K. and Meuth, S. G. (2018) 'The use of digital and remote communication technologies as a tool for multiple sclerosis management: narrative review', *JMIR rehabilitation and assistive technologies*, 5(1), pp. e5.
- Mattei, T. A. (2017) 'Privacy, confidentiality, and security of health care information: Lessons from the recent Wannacry Cyberattack', *World neurosurgery*, 104, pp. 972-974.
- McGraw, G. (2013) 'Cyber war is inevitable (unless we build security in)', *Journal of Strategic Studies*, 36(1), pp. 109-119.
- Morse, A. (2018) 'Investigation: WannaCry cyber attack and the NHS', *Report by the National Audit Office*. Accessed, 1.
- Ossebaard, H. C. and Van Gemert-Pijnen, L. (2016) 'eHealth and quality in health care: implementation time', *International journal for quality in health care*, 28(3), pp. 415-419.
- Palca, J. and Myre, G. (2020) *US, Canada, Britain say Russian Hackers Are After COVID-19 Vaccine Data*. Available at: <https://www.npr.org/2020/07/17/892195706/u-s-canada-britain-say-russian-hackers-are-after-covid-19-vaccine-data?t=1610022984774>.
- Romanosky, S. and Goldman, Z. (2016) 'Cyber collateral damage', *Procedia Computer Science*, 95(2), pp. 10-17.
- Ross, J., Stevenson, F., Lau, R. and Murray, E. (2016) 'Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update)', *Implementation science*, 11(1), pp. 146.
- Sevis, K. N. and Seker, E. 'Cyber warfare: terms, issues, laws and controversies'. 2016: IEEE, 1-9.
- Tidy, J. (2020) *Police Launch Homicide Inquiry After German Hospital Hack*. Available at: <https://www.bbc.com/news/technology-54204356>.
- Walker-Roberts, S., Hammoudeh, M. and Dehghantanha, A. (2018) 'A systematic review of the availability and efficacy of countermeasures to internal threats in healthcare critical infrastructure', *IEEE Access*, 6, pp. 25167-25177.
- Wickström, G., Regner, Å. and Micko, L. (2017) 'Vision eHealth 2025 common starting points for digitization in social services and health and medical care', *Affairs*. Available online: <https://www.ehalsomyndigheten.se/globalassets/dokument/vision/vision-for-ehealth-2025.pdf> (accessed on 10 December 2019).
- Williams, P. A. H. and Woodward, A. J. (2015) 'Cybersecurity vulnerabilities in medical devices: a complex environment and multifaceted problem', *Medical Devices (Auckland, NZ)*, 8, pp. 305.
- Wosik, J., Fudim, M., Cameron, B., Gellad, Z. F., Cho, A., Phinney, D., Curtis, S., Roman, M., Poon, E. G. and Ferranti, J. (2020) 'Telehealth Transformation: COVID-19 and the rise of Virtual Care', *Journal of the American Medical Informatics Association*, 27(6), pp. 957-962.
- Zhang, A., Wang, L., Ye, X. and Lin, X. (2016) 'Light-weight and robust security-aware D2D-assist data transmission protocol for mobile-health systems', *IEEE Transactions on Information Forensics and Security*, 12(3), pp. 662-675.
- Zubair, M., Unal, D., Al-Ali, A. and Shikfa, A. 'Exploiting bluetooth vulnerabilities in e-health IoT devices'. 2019, 1-7.
- Zubaydi, F., Saleh, A., Aloul, F. and Sagahyroon, A. 'Security of mobile health (mHealth) systems'. 2015: IEEE, 1-5.

# Talos: A Prototype Intrusion Detection and Prevention System for Profiling Ransomware Behaviour

Ashley Charles Wood, Thaddeus Eze and Lee Speakman

University of Chester, UK

[ashley.wood@chester.ac.uk](mailto:ashley.wood@chester.ac.uk)

[t.eze@chester.ac.uk](mailto:t.eze@chester.ac.uk)

[l.speakman@chester.ac.uk](mailto:l.speakman@chester.ac.uk)

DOI: 10.34190/EWS.21.026

**Abstract:** In this paper, we profile the behaviour and functionality of multiple recent variants of WannaCry and CrySiS/Dharma, through static and dynamic malware analysis. We then analyse and detail the commonly occurring behavioural features of ransomware. These features are utilised to develop a prototype Intrusion Detection and Prevention System (IDPS) named Talos, which comprises of several detection mechanisms/components. Benchmarking is later performed to test and validate the performance of the proposed Talos IDPS system and the results discussed in detail. It is established that the Talos system can successfully detect all ransomware variants tested, in an average of 1.7 seconds and instigate remedial action in a timely manner following first detection. The paper concludes with a summarisation of our main findings and discussion of potential future works which may be carried out to allow the effective detection and prevention of ransomware on systems and networks.

**Keywords:** IDS, IPS, IDPS, ransomware, WannaCry, CrySiS/Dharma

---

## 1. Introduction

### 1.1 Introduction

In our previous paper (Wood & Eze, 2020), we examined the way in which ransomware interacts with the system on infection to implicate upon both data and system functionality. Our key finding was that it was possible to restore data and system functionality following ransomware infection. This paper iterates on our previous work (Wood & Eze, 2020) and explores the prospect of profiling the behaviour of ransomware and developing an Intrusion Detection and Prevention System (IDPS) system based exclusively on the commonly occurring system behaviours of ransomware. Section two provides an overview of ransomware in recent times, section three summarises previous research in this area, section four summarises our behavioural analysis of WannaCry and CrySiS/Dharma, section five outlines and details the proposed Talos system and its detection performance. Section six concludes the paper with a summary and discussion of this study's main findings, before drawing the paper to a close with an overview of areas requiring further work to advance the state-of-the-art in IDPS technology.

### 1.2 Relevant terminologies

This paper refers to several acronyms and terminologies throughout, these are Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), Intrusion Detection and Prevention Systems (IDPS) and Ransomware. Firstly, Intrusion Detection Systems (IDS) automate the process of manual intrusion detection processes by monitoring networks and systems for malicious/suspicious activity in violation of established security policies. If activity is identified, activity is logged and alerts sent to administrators (Azhagiri *et al*, 2015). Comparatively, Intrusion Prevention Systems (IPS) share the same capability of an IDS but additionally respond to and take remedial action when malicious activity is identified, before notifying administrators of the activity detected and remedial action taken (Azhagiri *et al*, 2015). An Intrusion Detection and Prevention System (IDPS) as the name implies combines the capabilities of both IDS and IPS to formulate a more robust system. Ransomware refers to a type of malicious software which is designed to restrict access to a computer system and its data until such a time a monetary fee is paid.

## 2. Background

As technology evolves to become more advanced and sophisticated, so have the attackers, who are continually designing and developing ever more destructive and imaginative means of breaching network and system security. As society, the economy and critical infrastructure, increasingly depend upon information technology (IT), cyberattacks are becoming increasingly attractive to attackers with potentially disastrous consequences

(Jang-Jaccard & Nepal, 2014). The COVID pandemic has exacerbated the growing issue of malware/ransomware attacks, due to organisations swiftly adapting business infrastructures, which has left multiple loopholes within IT systems, presenting attackers with easy opportunities for exploitation (Check Point, 2020). As of Q3 2020, a 50% average daily increase of attacks globally was observed, compared to the first half of 2020, specifically, the USA saw a 98.1% increase, India a 39.2% increase, Sri Lanka a 43.6% increase, Russia a 57.9% increase and Turkey a 32.5% increase (Check Point, 2020). Furthermore, 90% of security professionals report growing volumes of cyberattacks over the previous 12 months in 2020, with 4/5 reporting attacks are growing more sophisticated than 2019 (Bannister, 2020).

Ransomware remains a prevalent cybersecurity threat, capable of causing serious disruption globally. In 2017, the WannaCry ransomware affected the United Kingdom's NHS and spread rapidly across NHS networks, causing unprecedented disruption and resulting in an inability to provide patient care, with 34 trusts locked out of devices, and 46 reporting interruption (Smart, 2018). This incurred costs of £92,000,000 and resulted in 19,000 appointments being cancelled (Goud, 2018).

Petya, also caused substantial disruption in 2017. Merck Pharmaceuticals experienced disruption to research, manufacture, and product sales, amounting to damages of \$670,000,000 (Davis, 2017). Whilst shipping companies, Maersk and TNT suffered substantial interference to operations and incurred nine figure costs (Greenberg, 2018). To recover, Maersk needed to reinstall 4000 servers, 45,000 systems and 2,500 applications over a 10-day recovery operation, which Maersk warns could incur losses amounting to \$300 million due to severe disruption (Osborne, 2018).

Another variant, CrySiS/Dharma, has become increasingly active recently, with activity increasing 148% from February to April 2019 (Arntz, 2019). Throughout 2018, new CrySiS/Dharma variants were discovered from January to August 2018 with further increases from September to November 2018 (Coveware, 2018). If payment is made, chances of decryption range from 25% to 100%, due to some variants being more sophisticated than others (Coveware, 2018). Evidence suggests CrySiS/Dharma is becoming more sophisticated with perpetrators quashing issues which allowed decryption without payment (Nadeeau, 2018), suggesting active interest in developing increasingly destructive ransomware.

### **3. Previous work**

Previous studies have explored the prospect of building an IDPS for the detection and prevention of ransomware. Firstly, Azer & El-Kosiary (2018), proposed an IDPS model for detecting network intrusions and ransomware, which builds upon the concept of honeypots. Multiple decoy files are placed on the system in areas not ordinarily accessed by legitimate users, decoy files are then monitored for any access attempts. The proposed IDPS model is tested in its ability to detect a variety of ransomware types such as Cryptowall, Kovter, Winlock, Cryptolocker, Filecoder and Reveton. Samples of each ransomware are then executed on a system monitored by the IDPS model. This model detected all variants with the slowest detection time being Filecoder at 25 seconds, whilst the fastest was Cryptolocker at 15 seconds (Azer & El-Kosiary, 2018). The model is also capable of detecting attacker intrusions, using techniques such as decoy tokens, decoy servers, decoy partitions and decoy shared folders. The model upon testing, could detect all attacker intrusions, with the slowest detection time being 13 minutes whilst the fastest was 5 minutes (Azer & El-Kosiary, 2018). Evidently, whilst the system could detect all intrusions and ransomware, its detection times varied, meaning intrusions and ransomware threats are left momentarily uninterrupted, which is undesirable, hence further work is required to reduce detection times.

Celdrán *et al* (2019) developed a system intended to detect and prevent ransomware from spreading within Integrated Clinical Environments (ICE). The system utilises machine-learning (ML) techniques to detect and classify the propagation phase of ransomware attacks, whilst Network Function Visualisation (NFV) and Software Defined Network (SDN) paradigms are implemented to prevent ransomware from spreading by isolating and replacing infected network devices. Celdrán *et al* (2019) performed tests which showed the system can detect ransomware such as WannaCry, BadRabbit, Petya and PowerGhost. This is achieved by performing anomaly detection using techniques such as One-class Support Vector Machine (OC-SVM), Local Outlier Factor (LOF) and Isolation Forest (IF), whilst techniques such as; Neural Networks (NN), Naïve Bayes (NB) and Random Forest (RF) are used for classification (Celdrán *et al*, 2019). Tests are performed for each technique, with OC-SVM proving most effective for initial attack detection with 92.32% precision and 99.97% recall, whilst NB was most effective in botnet attack classification with 99.99% accuracy within 0.22 seconds (Celdrán *et al*, 2019).







To assess the effects on data, the encryption behaviour of the WannaCry samples was monitored with FolderChangesView. This indicates files are not encrypted directly, but rather encrypted duplicates of files created before the originals are deleted. This behaviour is shown with the file “ffc.pdf” where an encrypted duplicate is firstly created before the original is erased (Figure 6), indicating files are recoverable following alleged encryption as established in our previous paper (Wood & Eze, 2020).

Filename	Modified Count	Created Count	Deleted Count	Renamed Count	Full Path
3D3C4D.tmp	1	1	1	0	C:\Data Set\3D3C4D.tmp
ffc.pdf.WNCRY	1	1	0	1	C:\Data Set\ffc.pdf.WNCRY

Filename	Modified Count	Created Count	Deleted Count	Renamed Count	Full Path
TempWebSiteWeb20...	3	1	1	0	C:\Users\user\AppDataL...
ffc.pdf	0	0	1	0	C:\Data Set\ffc.pdf

Figure 6: Encrypted duplicate of “ffc.pdf” created and original later deleted

Regarding files/data, static analysis with PEiD indicated WannaCry modifies file access permissions by processing the icacls command (Figure 7). If utilised within the C:\ directory, every user would have access to all files (Plett & Poggemeyer, 2017), under which WannaCry runs as a process, posing serious implications for file integrity.

```
0000F4FC 0000F4FC icacls . /grant Everyone:F /T /C /Q
```

Figure 7: References to icacls

The second sample revealed notable network activity, specifically FakeNet-NG indicated the sample during execution attempts to connect to a unknown domain (Figure 8). Which, Newman (2017) argues, acts as a kill switch. Furthermore, ARP protocol traffic is observable during analysis with Wireshark (Figure 9), this behaviour is exhibited by all samples except for the first, analysis revealed such behaviour exists to find other potentially vulnerable hosts on the network to infect.

```
03/06/18 02:11:37 PM [Diverter] pid: 924 name: WannaCry2.exe
03/06/18 02:11:37 PM [DNS Server] Received a request for domain 'www.fff
arfcndp91f3apoudfj0gozur1jfamurqurpwa.com'.
03/06/18 02:11:37 PM [DNS Server] Responding with '192.0.2.123'
```

Figure 8: WannaCry requests kill switch domain

Time	Source IP	Destination IP	Protocol	Details
6738	200.100.75	Vmware_Bd1e:78	Broadcast	ARP: 43 who has 100.254.100.97 Tell 100.254.4.66
6739	200.100.741	Vmware_Bd1e:78	Broadcast	ARP: 43 who has 100.254.100.97 Tell 100.254.4.66
6748	200.100.754	Vmware_Bd1e:78	Broadcast	ARP: 42 who has 100.254.100.97 Tell 100.254.4.66
6741	200.100.771	Vmware_Bd1e:78	Broadcast	ARP: 42 who has 100.254.100.97 Tell 100.254.4.66
6742	200.100.709	Vmware_Bd1e:78	Broadcast	ARP: 42 who has 100.254.100.97 Tell 100.254.4.66

Figure 9: ARP activity from WannaCry observed within Wireshark

Further static analysis of the first sample was carried out with Strings. This revealed references to 3 specific directories and 177 filetypes (Figure 10). Analysis indicated; these are the filetypes that WannaCry encrypts whilst the referenced directories appear to be excluded from the encryption process.

```
.vsd - .ppsm
.edh - .ppxc
.enl - .ppzm
.mcg - .ppz
.oat - .hul
.pst - .ppsm %s\%s
.potn - .ppxc %s\Intel
.potx - .ppt %s\ProgramData
```

Figure 10: Filetypes and directories referenced within first WannaCry sample

### 4.3 CrySiS/Dharma analysis

CrySiS/Dharma was also analysed, which upon execution will request administrator privileges, if granted, this results in immediate encryption of network drives (Figure 11), before data with the “C:\” directory and its subdirectories are encrypted (Figure 12). All variants create the “FILES ENCRYPTED.txt” file, containing a ransom/instruction note with alternating contact addresses in multiple locations. Finally, an interface is displayed demanding a ransom (Figure 13).

Filename	Modified	Created	Deleted	Renamed	Full Path
sample.mv.id-00000000 (locky)...	5/5/2018 4:00 PM	5/5/2018 4:00 PM	0	0	COMBO File 400 KB
sample.mv.id-00000000 (locky)...	5/5/2018 4:00 PM	5/5/2018 4:00 PM	0	0	COMBO File 400 KB
sample.mv1.id-00000000 (locky)...	5/5/2018 4:00 PM	5/5/2018 4:00 PM	0	0	COMBO File 35 KB
sample.mv4.id-00000000 (locky)...	5/5/2018 4:00 PM	5/5/2018 4:00 PM	0	0	COMBO File 375 KB
sample.mvq.id-00000000 (locky)...	5/5/2018 4:00 PM	5/5/2018 4:00 PM	0	0	COMBO File 407 KB

Figure 11: Mapped network drive contents encrypted

Filename	Modified	Created	Deleted	Renamed	Full Path
chrome.exe.id-00000000 (locky)...	4/10/2018 12:11 PM	4/10/2018 12:11 PM	0	0	COMBO File 2,304 KB
chrome.VisualElementsManifest.xml.id-0...	4/10/2018 12:11 PM	4/10/2018 12:11 PM	0	0	COMBO File 1 KB
master_preferences.id-00000000 (locky)...	4/10/2018 12:11 PM	4/10/2018 12:11 PM	0	0	COMBO File 125 KB



## 5. Proposed Talos IDPS system

### 5.1 Introduction

In this section of the paper, we present the prototype Talos IDPS system and discuss some of the ransomware behavioural features selected from the earlier ransomware analysis and components which have been developed to construct Talos. The naming of the Talos prototype system, is taken from Greek mythology and is named after the giant automaton who defended Europa in Crete from pirates and invaders by circling the island three times daily and hurling boulders at approaching enemy ships.

### 5.2 Common features of ransomware

After ascertaining the common behaviours of the ransomware samples, it was evident, many indicators of a ransomware infection manifested within the windows filesystem and as system processes. Thus, as a starting point for Talos, a selection of key filesystem features for both ransomware families are selected (Table 2). Notably, other behaviours, specifically file integrity issues may be generalisable to other ransomware variants, this aspect however will be explored as part of our future work.

**Table 2:** WannaCry and CrySiS/Dharma sample common features

<b>WannaCry</b>	<b>File Types</b>	.wnry		.wncryt		.wncry	
	<b>Processes</b>	tasksche.exe		taskse.exe		taskdl.exe	
	<b>Created Files</b>	C:\Users\User\Desktop\@WanaDecryptor@.bmp			C:\@Please_Read_Me@.txt		
		C:\Users\User\Desktop\@WanaDecryptor@.exe			C:\@WanaDecryptor@.exe		
C:\Users\User\Desktop\@Please_Read_Me@.txt			C:\Windows\tasksche.exe				
<b>CrySiS/Dharma</b>	<b>File Types</b>	.combo	.HARMA	.PLEX	.2020	.aa1	
	<b>Processes</b>	mshta.exe					
	<b>Created Files</b>	C:\Users\User\Desktop\FILES ENCRYPTED.txt			C:\Windows\System32\Info.hta		
		C:\Users\User\AppData\Roaming\Info.hta			C:\FILES ENCRYPTED.txt		

### 5.3 Talos prototype components

Considering information ascertained during analysis, and to allow development of Talos, it was determined several key components would be required, which included a main component, a file integrity check, a filetype check, a blacklisted file check, process check and reset/initialisation components. Each component was created in Python, the functions of each are explained in Table 3.

**Table 3:** Descriptions of each component of Talos

Component	Description
Main	The main component sequentially calls and executes each component and interprets the output of each i.e., status codes and takes remedial action where required. This is achieved by a function known as the “watcher” and the use of a centralised concern level variable, which is raised in the event activity is detected. If the variable raises to 1, the function disables all network adapters to prevent ransomware propagation, whereas if raising to 2 or above, the main component ensures all network adapters are disabled and the system is safely shut down to eliminate the risk of further damage. The main component will also call the initialisation and reset components if honey file integrity is reported as damaged by the “File Integrity Check” component.
Initialisation	The initialisation component, when called, generates multiple “.txt” honey files across various system directories, where ransomware activity is known to occur, such as the; “C:\”, user’s directory amongst other strategic locations. Such files contain a series of ASCII characters and digits of randomised lengths, to evade detection by malicious software. Upon generation, SHA256 and the paths/locations for each file are saved, to allow later recall and integrity checks by the File Integrity Check component.
File Integrity Check	Ransomware commonly affects file integrity, either through deletion or modification. This component verifies file integrity by gathering a list of files generated earlier by the initialisation component and their SHA256 hashes, it then generates new SHA256 hash for each of the files and compares them against the earlier generated SHA256 hashes. If these hashes match, then no file integrity issues have occurred, whereas if it does not match or files no longer exist, this strongly indicates file integrity is impacted, in this instance the component will raise a status code for remediation by the main component.
Filetype Check	This component examines defined directories for filetypes associated with ransomware, to achieve this, a list of files within each directory is gathered, and checks performed to verify if any end with a defined ransomware extension. If found, file details are logged and a status code raised, to allow later remediation.

Component	Description
Process Check	This component gathers a list of running system processes and checks this list to establish if it contains the name of an associated ransomware process loaded from a file. If a process is found to be running, the component will generate a log and raise an alarm for interpretation by the main component.
Blacklisted File Check	During execution, ransomware creates multiple files in various locations across the system, containing payloads or other required data. This component examines various defined directories for the presence of such files. If files are found, the component raises an alarm for remediation by the main component.
Reset	If the file integrity check reports that a generated honey file is damaged, then the reset component is called by the main component to erase all previously generated files before the initialisation component is called to generate a new selection of honey files for further integrity checks to be performed.

## 5.4 Data

One of Talos’s key features is the consideration of known ransomware behaviours to detect activity, to permit this, data on ransomware behaviour is stored within files. This data includes the list of files/ hashes generated by the initialisation component, the processes, files and filetypes associated with ransomware and the resulting status codes of each Talos component amongst other data. Each data item is stored within XML tags (Figure 20), which is later retrieved through regular expression parsing by each component. This method of storing and retrieving data, and the modularity of Talos will permit later adaptation of the system to allow the detection of further malware/ransomware variants.

```

<RANSOMWAREEXECUTABLE>tasksche.exe</RANSOMWAREEXECUTABLE>
<RANSOMWAREEXECUTABLE>@nsaDecrypton9.exe</RANSOMWAREEXECUTABLE>
<RANSOMWAREEXECUTABLE>taskd1.exe</RANSOMWAREEXECUTABLE>
<RANSOMWAREEXECUTABLE>taskise.exe</RANSOMWAREEXECUTABLE>
<RANSOMWAREEXECUTABLE>mshta.exe</RANSOMWAREEXECUTABLE>
<screeningpath>C:\</screeningpath>
<screeningpath>C:\Users\</screeningpath>
<screeningpath>C:\Users\KTALOSUSERS\</screeningpath>
<screeningpath>C:\Users\KTALOSUSERS\Documents\</screeningpath>
<screeningpath>C:\Users\KTALOSUSERS\Desktop\</screeningpath>
<screeningpath>C:\Users\KTALOSUSERS\Pictures\</screeningpath>
    
```

Figure 20: Two example of files from where components will read data from

## 5.5 Prototype system

Upon compiling data on common ransomware behaviours and establishing the required detection mechanisms, all required components of Talos could be developed. Once the system was built; it could then be executed as a console application (Figure 21). Talos executes each component sequentially as part of the main loop and interprets the results at each stage for action to be taken where necessary. If a component detects a single event, then a centralised “concernlevel” variable is incremented by “1”, whereas if a component reports multiple events, it is set to “2”. A function known as the “watcher” will check the value of this variable after executing each component, if it reaches “1”, all network adapters are disabled to minimise the spread of ransomware, whereas if greater than or equal to “2”, all network adapters are disabled and the system safely powered off to prevent further damage, as shown within the code snippet (Figure 22).

```

0-2038 - TALOS IDPS: Prototype, All Rights Reserved, Ashley Wood
0-2038 - TRYING INITIALISATION STATE
0-2038 - SYSTEM ALREADY INITIALISED, CONTINUING...
0-2038 - NO SUSPICIOUS ACTIVITY REPORTED
0-2038 - NO RANSOMWARE FILE TYPES DETECTED
0-2038 - NO RANSOMWARE ASSOCIATED FILE DETECTED
0-2038 - NO RANSOMWARE PROCESSES DETECTED
    
```

Figure 21: Main component successfully running all components

```

def watcher():
    global concernlevel, concernleveloneprocess
    if concernlevel == 0:
        os.system('color 7')
    if concernlevel > 0:
        previousconcernlevel = concernlevel
        if concernlevel == 1:
            if concernleveloneprocess == 0:
                concernleveloneprocess=1
                requestnetworkdisable=1
                networkdisable()
            os.system('color 6')
        if concernlevel >= 2:
            os.system('color 4')
            if concernleveltwoprocess == 0:
                concernleveltwoprocess=1
                requestnetworkdisable=1
                networkdisable()
                requestpoweroff=1
                systempoweroff()
    
```

Figure 22: The “watcher” function (source code)

In addition to performing remedial action, each component of Talos will create respective logs of the detected activity, for example the File Integrity check component, will record details of the activity detected, the expected and generated SHA256 hashes and additionally details of the current user, IP address, running processes and listening network services (Figure 23). All of which are collected for examination at a later point in time to ascertain precisely what occurred on the system to cause the activity.

```

EVENT DETECTED AT: 08-23-2020 07:00:29
PATH OF FILE INTEGRITY CHECKED: C:\sm07h.txt
EVENT TYPE: FILE NOT FOUND AT PATH LOCATION
EXPECTED SHA256: 99f24447ba8f88f045361eba79af70fe7017ac191c6239680c925e5815c80ad194
USER: Ashley
HOSTNAME: ASHLEY-ASUS-II
HOST ADDRESS: 192.168.1.130
LISTENING SERVICES:
TCP 0.0.0.0:135 0.0.0.0: LISTENING 1868
TCP 0.0.0.0:445 0.0.0.0: LISTENING 4
TCP 0.0.0.0:5000 0.0.0.0: LISTENING 6088
TCP 0.0.0.0:5557 0.0.0.0: LISTENING 4
TCP 0.0.0.0:38383 0.0.0.0: LISTENING 9552
TCP 0.0.0.0:45082 0.0.0.0: LISTENING 8884
TCP 0.0.0.0:45633 0.0.0.0: LISTENING 8884
TCP 0.0.0.0:49664 0.0.0.0: LISTENING 912
TCP 0.0.0.0:49665 0.0.0.0: LISTENING 756
TCP 0.0.0.0:49666 0.0.0.0: LISTENING 1524
TCP 0.0.0.0:49667 0.0.0.0: LISTENING 2428
TCP 0.0.0.0:49668 0.0.0.0: LISTENING 3268
TCP 0.0.0.0:49671 0.0.0.0: LISTENING 888
TCP 0.0.0.0:49928 0.0.0.0: LISTENING 8008
TCP 127.0.0.1:5354 0.0.0.0: LISTENING 4552
TCP 127.0.0.1:27925 0.0.0.0: LISTENING 4888
TCP 192.168.1.130:130 0.0.0.0: LISTENING 4
TCP [::]:135 [::]:0 LISTENING 1868
TCP [::]:445 [::]:0 LISTENING 4
TCP [::]:5357 [::]:0 LISTENING 4
TCP [::]:45401 [::]:0 LISTENING 9552
TCP [::]:45633 [::]:0 LISTENING 8884
TCP [::]:49664 [::]:0 LISTENING 912
TCP [::]:49665 [::]:0 LISTENING 756
TCP [::]:49666 [::]:0 LISTENING 1524
TCP [::]:49667 [::]:0 LISTENING 2428
PROCESSES RUNNING AT TIME OF INCIDENT:
Image Name PID Session Name Session# Mem Usage
System Idle Process 0 Services 0 0 K
System 4 Services 0 1,744 K
Registry 104 Services 0 96,530 K
smc.exe 420 Services 0 1,236 K
csrss.exe 620 Services 0 5,536 K
wininit.exe 756 Services 0 8,848 K
services.exe 900 Services 0 10,856 K
lsass.exe 916 Services 0 36,816 K
svchost.exe 952 Services 0 3,928 K
svchost.exe 548 Services 0 26,608 K
fontsubst.exe 592 Services 0 3,440 K
WDFHost.exe 768 Services 0 91,652 K
svchost.exe 1080 Services 0 18,280 K
svchost.exe 1090 Services 0 8,576 K
svchost.exe 1252 Services 0 6,844 K
svchost.exe 1356 Services 0 13,784 K
svchost.exe 1368 Services 0 11,620 K
svchost.exe 1572 Services 0 8,456 K
svchost.exe 1380 Services 0 9,884 K
svchost.exe 1388 Services 0 6,782 K
svchost.exe 1528 Services 0 12,180 K
svchost.exe 1536 Services 0 5,484 K
svchost.exe 1544 Services 0 10,740 K
svchost.exe 1700 Services 0 18,148 K
svchost.exe 1720 Services 0 6,892 K
svchost.exe 1780 Services 0 6,492 K
svchost.exe 1888 Services 0 15,424 K
svchost.exe 1884 Services 0 12,420 K
svchost.exe 1236 Services 0 8,836 K
svchost.exe 1488 Services 0 9,192 K
svchost.exe 2056 Services 0 11,180 K
    
```

Figure 23: Example of log file generated by the File Integrity Check component

### 5.6 Talos prototype performance benchmarking

After building the prototype and verifying its functionality, benchmarking was performed to measure the systems detection and response times to threats, specifically the earlier analysed WannaCry and CrySiS/Dharma samples. To perform benchmarking, a testing script was prepared, which; executes each ransomware sample, launches Talos and records the execution times of each. Benchmarking results indicated Talos could detect all variants of WannaCry and CrySiS/Dharma promptly (Table 4), with an average first detection time of 2 seconds for WannaCry and 1.6 seconds for CrySiS/Dharma, resulting in an average first detection time of 1.7 seconds. Results also indicate Talos can initiate remedial action within a reasonable timeframe with CrySiS/Dharma, although there is evidently a need to reduce these times with WannaCry.

Table 4: WannaCry and CrySiS/Dharma sample benchmarking results

WannaCry benchmarking results	Sample	Execution Time	First Detection	NWAD	SPO
	1	11:54:37	11:54:38	11:54:47	11:54:48
2	12:03:01	12:03:04	12:03:12	12:03:20	
3	11:43:59	11:44:01	11:44:09	11:44:10	
4	11:31:17	11:31:19	11:31:27	11:31:29	
CrySiS/Dharma benchmarking results	1	12:17:09	12:17:11	12:17:13	12:17:13
	2	12:27:01	12:27:03	12:27:05	12:27:05
	3	12:34:10	12:34:12	12:34:14	12:34:14
	4	12:40:03	12:40:04	12:40:07	12:40:07
	5	12:56:23	12:56:24	12:56:25	12:56:29

Notably across all CrySiS/Dharma variants with the exception of sample 5, the network adaptor disable (NWAD) and system power off (SPO) trigger times are identical, this occurred due to the file integrity check component detecting multiple incidents i.e. a file being modified and another deleted. Which immediately sets the components status code to a higher level, resulting in the main component setting the “concernlevel” variable to “2”, which results in the “watcher” function calling both the NWAD and SPO functions.

Benchmarking further revealed performance disparity between individual components, namely, the file integrity component proved most effective at detecting CrySiS/Dharma, whilst the blacklisted file check proved most effective at detecting WannaCry, made evident by the first detection order (Table 5). Where no result is recorded, Talos initiated remedial action before components detected activity. The ransomware associated filetype check proved least effective and only detected 1 WannaCry variant during testing. The performance disparity between each component is notable, as this indicates individual component performance is intrinsically linked to individual ransomware/malware behaviour, which suggests individual

components may prove more effective at detecting one variant over another. This finding may have potential ramifications if Talos is later adapted to account for other variants, and further suggests, combining multiple components may be required.

**Table 5:** Individual component detection when tested against WannaCry and CrySiS/Dharma variants

	Sample	File Integrity Check	Ransomware associated filetype check	Blacklisted File check	Ransomware process check
<b>WannaCry component detection results</b>	1	3		1	2
	2		1	2	3
	3	3		1	2
	4	3		1	2
<b>CrySiS/Dharma component detection results</b>	1	1			
	2	1			
	3	1			
	4	1			
	5	1		2	

## 6. Conclusions and future work

### 6.1 Study summary

In this study, the behaviour of multiple ransomware variants was profiled using static/dynamic analysis and later analysed to develop detection mechanisms for Talos, specifically focusing on filesystem activity. The system developed in this study, has shown an IDPS utilising the common behavioural features of ransomware/malware can prove highly beneficial in the active detection and mitigation of ransomware. The Talos system could detect all ransomware variants tested promptly, averaging 1.7 seconds for first detection.

Results achieved during performance benchmarking of Talos are promising, and represent an improvement over other comparable works, such as Azer & El-Kosairy’s (2018) study, where the detection time ranged from 15 seconds for Cryptolocker to 25 seconds for filecoder. Notably, Talos and the work of Azer & El-Kosairy (2018) are designed to detect different malware/ransomware types with the work of Azer & El-Kosairy’s (2018) able to detect other attack and intrusion types. Furthermore, Talos falls behind systems incorporating artificial intelligence-based techniques, such as the work of Celdrán *et al* (2019) where the classification time was quicker at 0.22 seconds. Consequently, further performance benchmarking and iteration of Talos is required to fully assess its performance against comparable systems. However, Talos is modular and may be later adapted to account for further malware/ransomware variants and other attack/intrusion types.

### 6.2 Future work

Whilst Talos offers a promising level of performance, there are areas which require further development to improve the accuracy and resiliency of the system. Firstly, Talos is at present designed and tested to detect several strains of CrySiS/Dharma and WannaCry, which constrains the system’s ability to detect other ransomware/malware variants and other intrusive/malicious activity. Thus, more complex network propagating ransoms such as Petya (Wood & Eze, 2020) and other forms of attacks/intrusions are not yet detected by Talos. Furthermore, the current approach taken with Talos, assumes that other security mechanisms such as firewalls and antivirus products have failed to contain threats to the point where ransomware/malware can attain a foothold on the system. To address these area, Talos will be further developed in our future work to incorporate the common behavioural characteristics of other intrusive activities and ransomware/malware variants and also to consider activity occurring on the wider network and filesystem as part of a hybrid host-based and network-based system. This will help to allow the earlier detection of threats and will drastically improve the detection capability of Talos, allowing it to detect a diverse range of attacks.

Whilst this study has specifically focused on CrySiS/Dharma and WannaCry behaviours, it is believed the behaviours uncovered may be generalised to other forms of ransomware i.e., the way in which ransomware creates encrypted duplicates of files before affecting the integrity of the originals. Other features, such as file creation, filetypes and process spawning may be generalisable to other variants, the system will however require further adaptation and testing further to account for this, this aspect will be addressed in our future work.



Another area requiring further develop is Talos's decision-making capabilities, present Talos performs two remedial actions in sequence, firstly disabling network adapters if one event is detected and secondly powering off the system if two or more events are detected. The aim of this is to prevent ransomware spreading to other systems and to prevent further damage to the system and data. This approach however is potentially problematic in the event of false-positive errors, where legitimate activity is erroneously perceived as malicious. This area is acknowledged as a problem, and will be addressed in our future work, to achieve this Artificial Intelligence (AI) will be implemented into Talos, to allow it make more informed decisions about the actions it takes by considering the characteristics of previously detected threats to determine how best to respond. AI and Machine-learning based techniques have received considerable interest in the wider-research community, Celdrán *et al* (2019) applied machine learning techniques to their proposed system and saw promising detection accuracy scores and classification times. William (2020) argues, AI-based IDS unlike traditional IDS, have capability to learn over time from previous attacks to allow creation of new detection algorithms, allowing it to learn how to stop stealthy adversaries. Our future work will aim to address these areas, to allow Talos to perform more effectively.

Whilst the performance of Talos is evidently promising, based upon the results of our own performance tests. It is acknowledged that Talos is a work in progress the rates of false-positives and false-negatives is not yet ascertained. At the present point of development, it is acknowledged that Talos could potentially perceive legitimate/benign activity as malicious and actively block it, or could fail to detect other forms of ransomware, outside of those analysed as part of this study, resulting in false-positive and false-negative errors. Our future work will measure these rates by simulating multiple benign and malicious activities on the system and recording the response of Talos to ascertain the true false-positive and false-negative error rates.

## References

- Arntz, P. (2019). *Threat spotlight: CrySIS, aka Dharma ransomware, causing a crisis for businesses*. Retrieved from <https://blog.malwarebytes.com/threat-analysis/2019/05/threat-spotlight-crysis-aka-dharma-ransomware-causing-a-crisis-for-businesses/>
- Azer, M, A., & El-Kosairy, A. (2018). Intrusion and Ransomware Detection System. 2018 1st International Conference on Computer Applications & Information Security (ICCAIS), 1-7. <https://doi.org/10.1109/CAIS.2018.8471688>
- Azhagiri, M., Karthik, S., & Rajesh, S. (2015). INTRUSION DETECTION AND PREVENTION SYSTEM: TECHNOLOGIES AND CHALLENGES. *International Journal of Applied Engineering Research*, 10(87), 1-12. [https://www.researchgate.net/publication/287208734\\_Intrusion\\_Detection\\_and\\_Prevention\\_System\\_Tchnologies\\_and\\_Challenges](https://www.researchgate.net/publication/287208734_Intrusion_Detection_and_Prevention_System_Tchnologies_and_Challenges)
- Bannister, A. (2020). *Remote working during coronavirus pandemic leads to rise in cyber-attacks, say security professionals*. Retrieved from <https://portswigger.net/daily-swig/remote-working-during-coronavirus-pandemic-leads-to-rise-in-cyber-attacks-say-security-professionals>
- Celdrán, A, H., Clemente, F, J, G., Gómez, A, L, P., Lee, I., & Weimer, J. (2019). Intelligent and Dynamic Ransomware Spread Detection and Mitigation in Integrated Clinical Environments. *Sensors* 2019, 19(5), 1-31. <https://doi.org/10.3390/s19051114>
- Checkpoint. (2020). *Global Surges in Ransomware Attacks*. Retrieved from <https://blog.checkpoint.com/2020/10/06/study-global-rise-in-ransomware-attacks/>
- Coveware. (2018). *Why New Dharma Ransomware is More Dangerous than ever*. Retrieved from <https://www.coveware.com/blog/2018/11/26/why-new-dharma-ransomware-is-more-dangerous-than-ever>
- Davis, J. (2017). *Petya attacks now appear to be causing permanent damage*. Retrieved from <https://www.healthcareitnews.com/news/petya-attacks-now-appear-be-causing-permanent-damage>
- Goud, N. (2018). *NHS lost £92 million and Cancelled 19K appointments due to WannaCry Ransomware Attack*. Retrieved from <https://www.cybersecurity-insiders.com/nhs-lost-92-million-and-cancelled-19k-appointments-due-to-wannacry-ransomware-attack/>
- Greenberg, A. (2017). *THE WANNACRY RANSOMWARE HACKERS MADE SOME REAL AMATEUR MISTAKES*. Retrieved from <https://www.wired.com/2017/05/wannacry-ransomware-hackers-made-real-amateur-mistakes/>
- Jang-Jaccard, J., & Nepal, S. (2014). A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences*, 80(5), 973-993. <https://doi.org/10.1016/j.jcss.2014.02.005>
- Microsoft. (2019). *davclnt.h header*. Retrieved from <https://docs.microsoft.com/en-us/windows/win32/api/davclnt/>
- Nadeau, M. (2018). *11 Ways Ransomware is Evolving*. Retrieved from <https://insights.samsung.com/2018/04/11/11-ways-ransomware-is-evolving/>
- Newman, L, H. (2017). *HOW AN ACCIDENTAL 'KILL SWITCH' SLOWED FRIDAY'S MASSIVE RANSOMWARE ATTACK*. Retrieved from <https://www.wired.com/2017/05/accidental-kill-switch-slowed-fridays-massive-ransomware-attack/>
- Osborne, C. (2018). *NonPetya ransomware forced Maersk to reinstall 4000 servers, 45000 PCs*. Retrieved from <https://www.zdnet.com/article/maersk-forced-to-reinstall-4000-servers-45000-pcs-due-to-notpetya-attack/>

**Ashley Charles Wood, Thaddeus Eze and Lee Speakman**

- Plett, C., & Poggemeyer, L. (2017). *Icacls*. Retrieved from <https://docs.microsoft.com/en-us/windows-server/administration/windows-commands/icacls>
- Smart, W. (2018). *Lessons learned review of the WannaCry Ransomware Cyber Attack*. Retrieved from <https://www.england.nhs.uk/wp-content/uploads/2018/02/lessons-learned-review-wannacry-ransomware-cyber-attack-cio-review.pdf>
- William, D. (2020). *How AI can help improve intrusion detection systems*. Retrieved from <https://gcn.com/articles/2020/04/15/ai-intrusion-detection.aspx>
- Wood, A. & Eze, T. (2020). The Evolution of Ransomware Variants. *Proceedings of the 19th European Conference on Cyber Warfare and Security ECCWS 2020* (pp. 410-420). Chester, United Kingdom: ACPI.



# **Masters Research Papers**



# The use of Neural Networks to Classify Malware Families

Theodore Drewes and Joel Coffman  
United States Air Force Academy, USA

[teddrew34@gmail.com](mailto:teddrew34@gmail.com)

[joel.coffman@usafa.edu](mailto:joel.coffman@usafa.edu)

DOI: 10.34190/EWS.21.060

**Abstract:** Many antivirus vendors detect and classify malicious software, but there is little consensus among vendors regarding the label assigned to each malware sample. With an increase in malware capability, new malware uses “automation to generate new variants of themselves” (Thanh and Zelinka 2019), creating relationships implying underlying families of malware to which individual samples of malware belong. In this work, we explore using a neural network to classify the family of a given malware sample, which is a first step to unify vendors’ labels into a single ground truth classification. A consistent taxonomy (i.e., classification scheme for malware) facilitates consistent communication regarding malware and improved malware detection. Experiments with a data set of 13,000 malware samples reveals the merits of our approach.

**Keywords:** malware classification, malware taxonomy, neural networks, machine learning

---

## 1. Introduction

Malicious software (i.e., malware) is one of the most dangerous, yet ill-defined, threats faced by our modern, technology-based society. A report from the Council of Economic Advisors to the White House estimated that malicious software costs the United States economy upwards of \$109 billion annually and projects the cost of such activity to only grow in the future (Council of Economic Advisors 2018). As the damages of malware increase, so do the protections needed against malicious code.

New malware samples are generated every day by evolving and obfuscating existing malware. Once a flaw in a system is exposed and exploited, other malware developers seek to exploit the same vulnerability until it is fixed. For example, when the Win32 API was first introduced, a common method for intrusion was overwriting the header (Szor 1998). Once this flaw was established, waves of malware targeted this vulnerability until being rendered ineffective by heuristic detection techniques. Attackers use the same process to compromise modern systems: identify an opening (i.e., a vulnerability), write malware to exploit that vulnerability, and create variants to avoid detection by antivirus vendors. Malware authors need not even be proficient in developing exploits or variants themselves. Exploit toolkits provide pre-written exploits to compromise a system (Cannell 2016), and virus construction kits allow even novices to harness polymorphic and metamorphic techniques to create unique variants of preexisting malware (Szor 2005). While the overall number of cyber-attacks decreased in 2020, over the first six months of 2020, SonicWall identified over 315,000 new malware variants, which is a 63% increase in the new malware discovered over any six-month period of the prior year (SonicWall 2020). Whether this increase in the detection of new malware is due to improved detection capability or a true increase in new malware variants, the sheer volume of existing malware poses a serious challenge to combat.

New malware variants are often derived from existing malware (de la Cuadra 2007; Thanh and Zelinka 2019) because writing malware from scratch is cost-prohibitive. The lineage between malware samples (i.e., the existing malware that serves as the basis for a new variant) defines distinct families of malware (Dumitras and Neamtiu 2011). For example, new variants may add functionality to an existing piece of malware, fix bugs that prevent the malware from successfully exploiting targets, or tweak their code in an attempt to evade signatures used by antivirus vendors to detect the malware. By formally defining which malware samples belong to a particular family, antivirus software can apply known strategies to combat malware when it is detected, and targeted countermeasures minimize the impact of protecting the system.

Unfortunately, antivirus vendors inconsistently label malware – i.e., identify the sample as belonging to a specific group or family of related samples. Each antivirus vendor has its own internal taxonomy, and even within vendors, related malware can be classified as belonging to different families. For example, Table 1 lists the labels assigned by several antivirus vendors to two malware samples, and there is only the most superficial similarity among the labels assigned by each vendor (e.g., the first sample is likely a Trojan horse associated with the “VilseI” family). Consistent labeling is beneficial because malware with similar features is often combatted in a

similar manner. Knowing a malware sample’s family can lead to a more effective defense. Moreover, consistent labelling facilitates information sharing across vendors.

**Table 1:** Sample of vendor classifications for two malware samples using data from VirusTotal

MD5 Hash	8c4a59f2e73ac5aaa7c7b9132f805260	b17f791d70f0a742331d82492127b590
Vendor A	W32.FamVT.ViselPM.Worm	Win32.HLLP.Shohdi
Vendor B	Trojan.GenericKD	Gen:Variant.Razy
Vendor C	Trojan.VilseI	Generic.mg

From an academic perspective, researchers implicitly depend upon consistent ground truth labels to evaluate malware detection schemes. Otherwise, evaluations may be subject to systematic bias as a result of comparing results to a single vendor (replicating that vendor’s bias – if any – in its classification) or against multiple vendors (in which case the inconsistency among vendors affects the evaluation metrics).

From an educational standpoint, the classification of malware samples into different families provides insight into the design and purpose of malware and also suggests how to design systems that are resilient to certain malware families. Furthermore, identifying malware into known families allows new malware samples that do not fit neatly into these families to be flagged for further study.

In this work, we explore the use of machine learning, specifically neural networks, to establish a consensus classification for malware samples. We design two families of neural networks that classify malware samples using data provided by antivirus vendors. Experiments using a data set of more than 13,000 malware samples reveals that our classification matches that provided by AVCLASS (Sebastián et al. 2016) 19.56% of the time, which places our work in the top 50% of antivirus vendors for the data set.

The remainder of this paper is organized as follows. Section 2 contains a broad overview of neural networks and summarizes the services provided by VirusTotal and AVCLASS. Section 3 describes the design and implementation of our neural networks. Section 4 evaluates our work using a malware data set with more than 13,000 samples. Section 5 summarizes related work. We conclude in Section 6, including potential extensions to our work.

## 2. Background

This section reviews background information about machine learning, specifically neural networks. We also describe VirusTotal and AVCLASS, two tools that we use for our experiments.

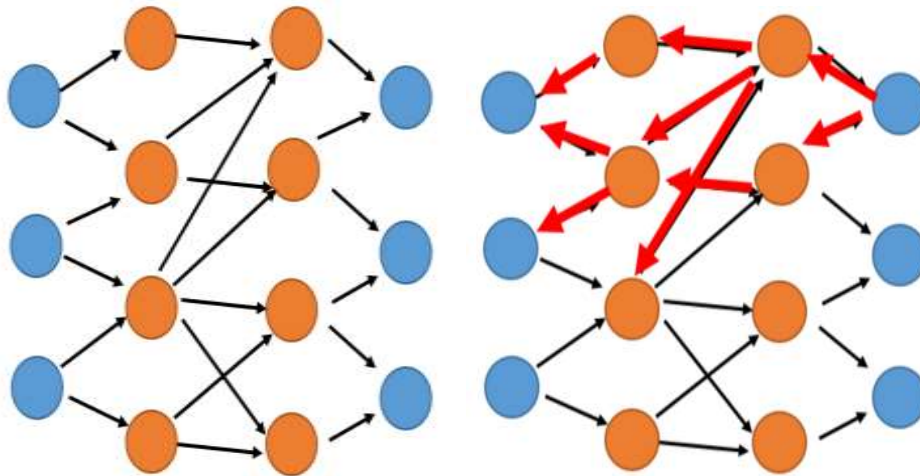
### 2.1 Machine learning and neural networks

Machine learning is the field of study in which computers are constructed to mimic human intelligence and to simulate learning. Such learning can be performed in many ways, but all techniques use a system of mathematical computation to “learn.” A subset of machine learning deals with artificial neural networks, often referred to as simply neural networks. These networks simulate the human process of decision making by utilizing artificial neurons arranged in layers to make decisions. Since first being introduced in the 1940s (McCulloch and Pitts 1943), neural networks have progressed to become a powerful tool in computer science (Clabaugh et al. 2000). They are used in a variety of fields from controlling nonplayable characters in video games to image recognition for the United States Department of Defense (Reddemann 2015).

A neural network takes in a data vector and returns a single output value as determined by the network. Classifying a malware sample as a belonging to a single family based on an input of vendor classifications fits within the set of problems that a neural network can solve. Much like a human recognizes features in an object and assesses what that object is, a neural network is given features to analyze in an input vector and determines what that input vector describes.

There are two primary ways for a neural network to pass information: forward propagation and backwards propagation. First, the neural network is provided an input array with each value corresponding to an input node. In forward propagation, the neural network passes these values forward from layer to layer, using a system of weights, biases, and node values to manipulate the value, ultimately arriving at the output nodes. Once at the output layer, the network identifies the output nodes with non-zero values; these nodes correspond to possible outputs, and the node with the highest value is typically chosen as the output of the neural network. The output (i.e., the computed answer) is compared to the ideal (i.e., correct) answer. In backwards propagation,

the weights are then updated in a backwards fashion to align with the ideal output more closely. Figure 1 illustrates both processes.



**Figure 1:** Illustrations of a forward propagating network (left) and backwards propagating network (right)

Evaluating a neural network uses three data sets. First, a training data set is used to set the weights and biases of nodes to maximize the accuracy of the network. Training uses forward propagation to pass an input vector through the network and backwards propagation to update the weights toward their ideal value. Second, a test data set is forward propagated through the network to evaluate the network’s performance. Third, the real-world data set is the data that the network encounters when deployed. Unlike the training and test data sets, real-world data does not have a known correct answer, and the developer of the network must be confident in the network’s ability to produce a correct answer. In theory, the test data set should closely resemble real-world data although parity tends to be difficult to achieve in practice.

## 2.2 VirusTotal and AVclass

Using a neural network to classify malware into families first requires defining the network’s inputs and outputs. In our work, antivirus vendors’ classifications of a malware sample provide the inputs. This data is available from VirusTotal,<sup>1</sup> an online tool that collects information on a given malware sample from over 70 vendors, including each vendor’s classification of the sample (see examples in Table 1). The output of the neural network is the family to which a given malware sample belongs.

Antivirus vendors classify malware according to their respective internal taxonomies; thus, the labels can differ from vendor to vendor as seen in Table 1. Some semblance of the true family can be inferred from manual analysis. For example, the first sample might be inferred to be a Trojan horse (Landwehr et al. 1994) belonging to the “Vilsel” family, which seeks to change the Windows firewall and download additional malware (MalwareBytes 2020). For an information system to use of the vendors’ classifications, additional work must be done to specify the family using a consistent taxonomy.

We use AVCLASS (Sebastián et al. 2016), one of the leading malware classification generators, as the ground truth due to the aforementioned lack of consistency among antivirus vendors. For example, AVCLASS labels the malware samples listed in Table 1 as “vilsel” and “shodi” respectively. While VirusTotal provides a list of vendors and classifications, AVCLASS provides a singular label for the malware family based on a consensus classification of the VirusTotal data. More details regarding AVCLASS appear in our review of related work (Section 5).

## 3. Using a neural network for malware classification

We consider two designs for our neural network for malware classification, specifically related to the number of input nodes. Because not every malware sample is classified by every vendor (i.e., a vendor may never have encountered – and consequently reported – a particular malware sample), most input nodes are not relevant to a particular malware sample. This remainder of this section describes both our neural network designs in detail.

<sup>1</sup> <https://www.virustotal.com/>

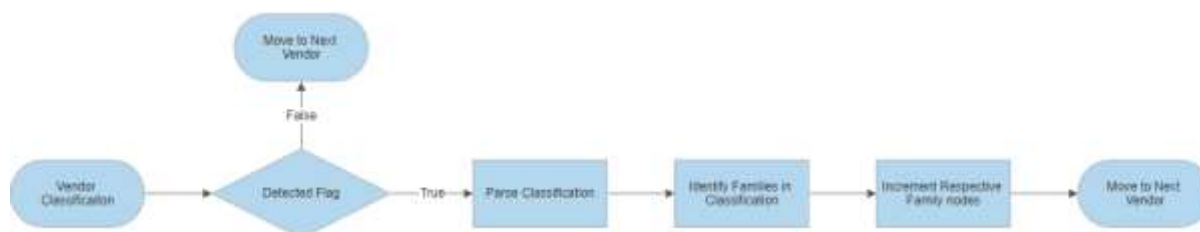
For both designs, we use a sigmoid activation function to keep node values between 0 and 1 across all layers of the neural network.

### 3.1 Generation 1

Vendor classifications are the input to our neural network. A vendor classification may have multiple components (e.g., “Trojan.Vilsel” in Table 1 can be decomposed into “Trojan” and “Vilsel”). We parse each vendor classification to create a set of labels. Our parser also maps known aliases to a canonical form (e.g., “tj” is mapped to “Trojan”). The weight assigned to each label is the reciprocal of the size of the set. For example, our parser assigns the label “Generic” a weight of 1.0 for “Generic.mg” (see Table 1); for “Trojan.Vilsel,” our parser assigns the label “Trojan” a weight of 0.5 and the label “Vilsel” a weight of 0.5. Thus, vendor classifications that correspond to multiple labels implicitly reflect uncertainty,<sup>2</sup> which we capture in the weights that we assign to each input node. Any input node corresponding to a label that is not contained in the vendor classification is set to 0.

Our first approach for the input values used a bit vector where each bit encoded a malware family. Thus, there is an input node for each vendor, and we interpret the bit vector as an integer when provided as input to the neural network. Unfortunately, this approach is problematic because the translation from a bit vector to an integer does not preserve the bit vector’s original semantics – in this case, labels assigned to high order bits are interpreted as being more significant than those assigned to low order bits in the representation. For example, 1100 0000 and 0000 0011 both represent sets that contain two elements, but the difference between 0100 0000 and 0000 0010 (sets that each contain one of the elements in the original sets) is significant when interpreted as an integer – 192 is much further from 64 than 3 is from 2. Consequently, we represent the vendor’s malware classification explicitly with a separate input node for each possible label (i.e., the bit vector becomes a series of binary inputs to the neural network). Another alternative is making each vendor’s classification an input node, which might provide more accurate results, but we leave this approach as an opportunity for future work.

Thus, our initial neural network design has an input node for each vendor and malware sample’s family label(s). Using our evaluation data set (see Section 4), there were 82 vendors and 908 unique labels; thus, there are 74,456 input nodes and 908 output nodes. AVCLASS labels malware identified by a single vendor as a singleton suffixed by the vendor’s label, and we group all these samples into a single classification. We use four hidden layers to gradually diminish the size of each layer at a rate that allows information to be condensed in a timely fashion, but not too quick to lose information. The respective rate of node decreases is approximately 0.125, 0.500, 0.500, 0.725, and 0.500, which is comparable to a well-performing network that we constructed for the MNIST data set (LeCun et al. 1998) as part of preliminary work.



**Figure 2:** Illustration of algorithm used to parse vendor data for a malware sample

Figure 2 depicts our algorithm for processing samples. First, we retrieve the VirusTotal data for the malware sample. For example, VirusTotal might report that a vendor detected a sample as malicious and assigned it the label “Trojan.GenericKD.Win32.87874.” We parse this vendor classification into three labels: Trojan, Generic, and Win32. The input node associated with each of these labels is assigned the value of 0.333..., and all other input nodes associated with the vendor are assigned the value of 0. This process is then repeated for other vendors.

<sup>2</sup> We assume that the vendor classification is not hierarchical (e.g., the Vilsel family of malware does not strictly comprise Trojan horses) although some antivirus vendors may use such a taxonomy. Unfortunately, vendors’ internal taxonomies are not consistent.

### 3.2 Generation 2

Initial experiments (see Section 4) indicated that the Generation 1 design suffered from poor performance due to the number of inputs and size of the neural network. Consequently, we designed a second generation of neural networks with several changes. First, the input layer only contains 907 nodes, one for each unique family that AVCLASS found in our evaluation dataset. For this design, we ignored all the singletons (i.e., malware identified by a single vendor), which accounts for the 907 vs. 908 input and output nodes. Second, the number of layers and nodes in each layer changes due to the decrease in the number of input nodes, and we constructed multiple neural networks with a variety of shapes for our evaluation.

Additionally, we modified the input vector for these neural networks, although we use the same approach to parsing vendor classifications. In these neural networks, we have an input node for each possible label and increment the counter of the input node each time that a vendor classifies a malware sample as belonging to the corresponding family. For example, if the label “Vilsel” is present in five vendors’ classifications, then the input node corresponding to “Vilsel” has a value of 5. If a potential family is not detected by any vendors, then the corresponding input node has a value of 0. This approach significantly reduces, but does not eliminate, the possibility for input values to be 0.

Because our Generation 2 neural networks have the same number of input and output nodes (e.g., 907 nodes each for our evaluation data set), we vary the number of nodes in the hidden layers, gradually increasing or decreasing the number of nodes and converging back on the number of output nodes. We also initially created a neural network with no hidden layers to evaluate the simplest possible network for this general approach.

## 4. Evaluation

We implemented our neural networks using Keras,<sup>3</sup> a Python library that provides a simplified interface to TensorFlow (Abadi et al. 2016). Major corporations across a variety of industries (e.g., NASA, the European Council for Nuclear Research (CERN), and the National Institutes of Health) all use Keras for neural networks. In preliminary experiments with the MNIST data set (LeCun et al. 1998), we found Keras balanced run time, error rate, and generality across problem domains. We trained and tested our neural networks on a Windows 10 machine with 4 processors and 16 GBs of RAM.

**Table 2:** Design of neural networks

Design	Description	Nodes		Layers	
		Input	Output	Total	Nodes per Layer
Generation 1	–	74456	908	6	74456, 10000, 5000, 2500, 1800, 908
Generation 2	No layers	907	907	2	907, 907
	Small	907	907	5	907, 600, 300, 600, 907
	Big	907	907	5	907, 1200, 1800, 1200, 907
	Big...Small	907	907	4	907, 1200, 600, 907
	Small...Big	907	907	4	907, 600, 1200, 907

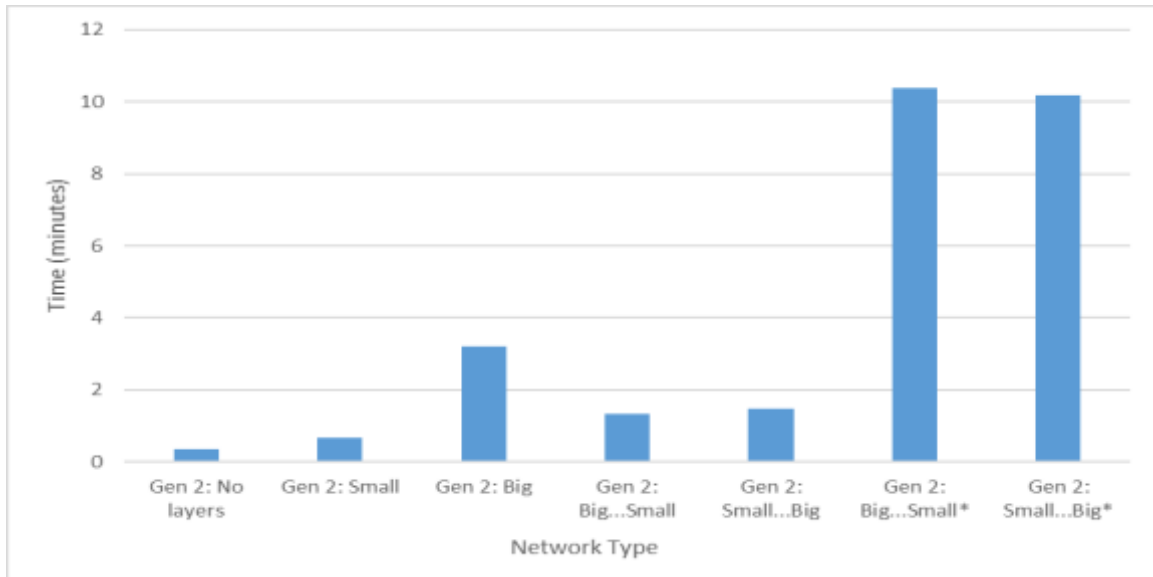
Our evaluation data set comprises 13,248 malware samples obtained from VirusTotal. The samples included malware that targeted the Win32 API, that were hidden in emails and websites, and that affected the Microsoft Office suite. VirusTotal reported classifications from 82 different vendors for this data set, but all the vendors did not classify each malware sample. Instead, the number of classifications for the samples ranged from 57 to 76 – i.e., no sample was classified by every vendor. Additionally, there were 908 unique AVCLASS labels for our evaluation data set, including a singleton when only a single vendor identified a sample as malicious. We use a 90/10 split for the training and test data sets for our experiments.

The Generation 1 neural network required more than 6 hours to run on the full dataset and produced a mean accuracy of 5.363% (standard deviation of 0.015%). A major limitation of this neural network is its input: the 908 malware families and 82 vendors resulted in 74,456 input values. This large number of inputs hindered the network’s ability to make an accurate assessment of a malware sample. In addition, the input values were sparse: if a vendor did not classify a specific malware sample, then there were 908 0s for that sample. Furthermore, if a vendor does identify a sample as malware it will not classify it as all 908 possible classifications, at most labeling it with a few of the possible families, resulting again in a high number of 0s for that sample.

<sup>3</sup> <https://keras.io/>

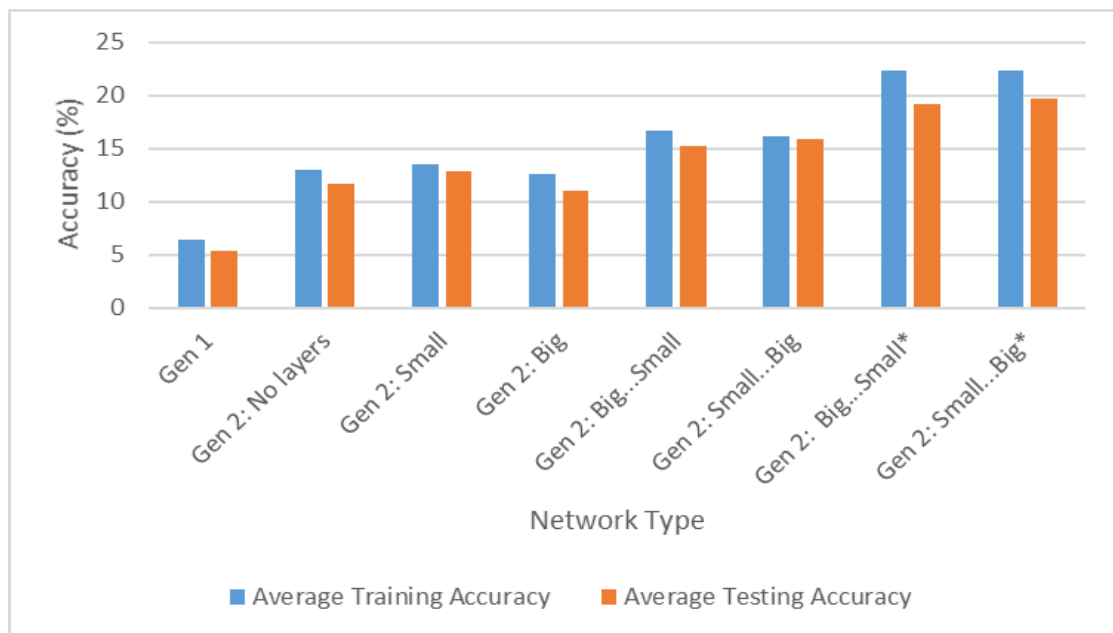
Because a majority of the samples produced an input array of mostly 0s, updating the weights in the network is spread over a large number of 0 nodes rather than the nodes that provided the meaningful input.

The lengthy runtime of the Generation 1 neural network is also attributable to the number of inputs coupled with the size of the network. For this network, there were  $9.863 \times 10^8$  pieces of input data (74,456 input nodes multiplied by 13,248 samples).



**Figure 3:** Accuracy of neural networks for training and test splits

In comparison, the Generation 2 neural networks are substantially faster and more accurate on average than the Generation 1 neural network. Over three runs for each neural network, the maximum test accuracy achieved was 19.64% by the “Small...Big” neural network (4 layers with 907, 600, 1200, and 907 nodes in the layers). See Figure 3 and Figure 4 for a comparison of the accuracy and run time of the various neural network designs.

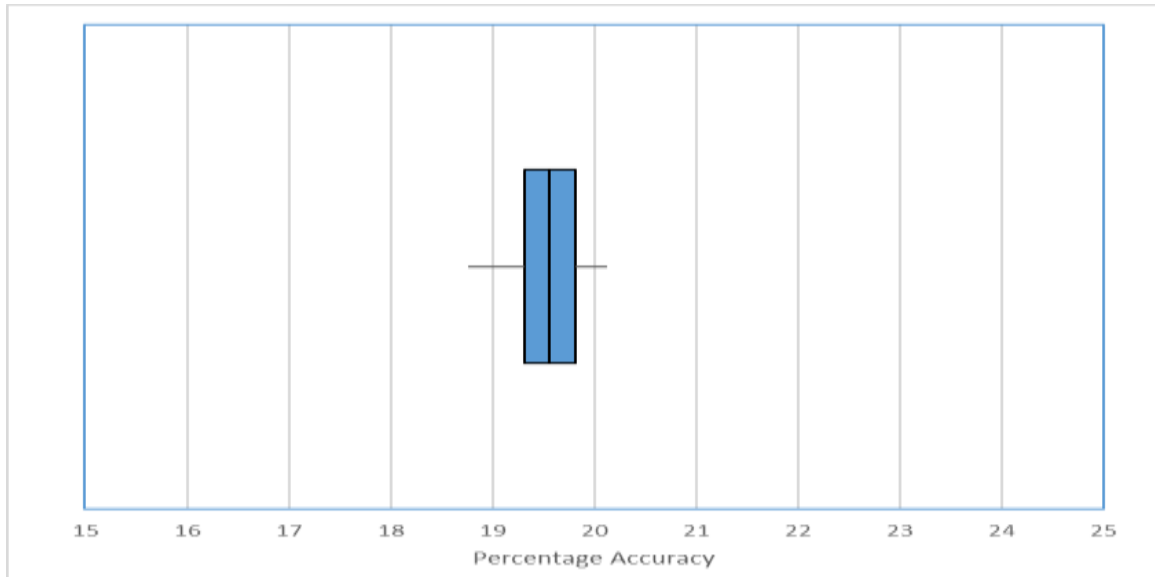


**Figure 4:** Mean run time of neural networks for three test runs

Of the Generation 2 neural networks, those that performed the best were the ones that varied the size of the hidden layers so that one hidden layer had fewer than 907 nodes and one hidden layer had more than 907 nodes. We selected these networks to experiment with the number of training epochs. While the original neural networks used 20 training epochs, we varied the number of training epochs up to a maximum of 300 training epochs. Although accuracy increased with additional training epochs, we saw a noticeable limit around 150



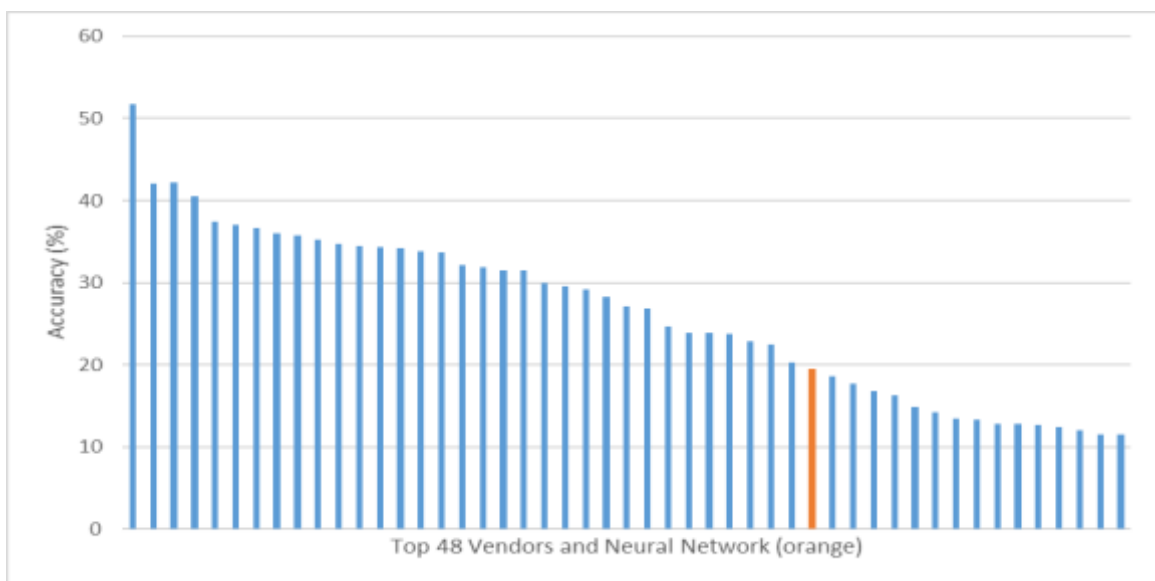
training epochs, which increased the accuracy by approximately 5%. For 150 training epochs, the “Small...Big” neural network (4 layers with 907, 600, 1200, and 907 nodes in the layers) again performed the best. The mean accuracy was 19.54% with a standard deviation of 0.38%. This network achieved a recall of 25.46%, a precision of 9.57%, and an F<sub>1</sub> score of 13.91.



**Figure 5:** Neural network performance over 25 iterations. The box identifies the middle quartiles, the center line identifies the median, and the whiskers extend to the minimum and maximum values

#### 4.1 Discussion

The improvement from our Generation 1 neural network to the Generation 2 neural networks is significant. The decrease in the number of input nodes improved the accuracy of the network. Because the input was condensed into 908 nodes, the number of possible input values that could be 0 was only 1.22% of the input values in the Generation 1 neural network. Consequently, the Generation 2 neural networks learned without the interference of as many “dead” input nodes in the training data. Additionally, the runtime of the Generation 2 neural networks decreased by two orders of magnitude due to the decrease in the overall size of the network. What originally required more than six hours was accomplished in a few minutes. While an accuracy of 19.54% may not seem impressive, our best network performs better than 49 of the 82 vendors in providing a label that matches with AVCLASS (see Figure 6).



**Figure 6:** Comparison of performance against the ground truth provided by AVclass. Our best-performing neural network is highlighted in orange, and its performance is in the top 50% of vendors

## 4.2 Threats to validity

The lack of consistent classifications for malware samples precipitated our work, yet our evaluation suffers from this exact issue. Without agreement by antivirus vendors to adhere to a consistent taxonomy, ground truth labels for malware samples are difficult to ascertain. Although we use AVCLASS as the baseline for our experiments, its simple plurality vote cannot capture the nuances of a well-designed taxonomy. Thus, even though our neural networks fail to reproduce AVCLASS's labels exactly, the deviations may actually indicate cases where AVCLASS's labels can be improved. We intend to investigate this possibility further as part of future work.

Our data set of 13,000 malware samples is small compared to the hundreds, if not thousands, of new malware discovered every day. Including additional data sets or a larger data set would minimize the risk of over-fitting our neural networks, ensuring that our results generalize.

## 5. Related work

In AVCLASS (Sebastián et al. 2016), researchers recognized that the antivirus engines that VirusTotal uses are inconsistent in labeling malware to a specific family. To standardize the family name for a specific piece of malware, the researchers used a plurality vote based on labels from the antivirus engines. This approach is similar to that taken by other researchers (Wei et al. 2017). Sebastián et al. addressed three main challenges of this method -- normalization, removal of generic tokens, and alias detection. After testing their solution on 10 malware data sets comprising over 8.9 million samples, the researchers achieved an F<sub>1</sub> score of 93.9 where the F<sub>1</sub> score is calculated using precision and recall, the correctly defined positives divided by all identified positives and the identified positives divided by all true positives respectively. While the goal of AVCLASS and our work is similar, we use machine learning, specifically a neural network, rather than a simple plurality vote where each vendor has equal weight. Thus, our approach allows for the inclusion of all a vendor's input into the final determination rather than only considering the most popular answer. Nevertheless, we use AVCLASS as the performance baseline for our work due to the lack of another data set with established ground truth.

Many powerful open source libraries exist for machine learning in general and neural networks in particular. The best known is TensorFlow,<sup>4</sup> a 2015 Google project with primary use in speech recognition, photograph identification and email autoreply (Abadi et al. 2016). In addition, there are several other open-source neural networks that lack the computing power of TensorFlow but are better in an individual metric including Caffe,<sup>5</sup> which focuses on the speed and efficiency of passing data through a network (Jia et al. 2014).

## 6. Conclusion

In this work, we evaluate the use of neural networks to consistently label malware. Consistently labeling malware samples improves cyber defense by eliminating ambiguity across antivirus vendors and by facilitating information exchange. We realize differences in antivirus vendors' internal detection and classification inhibit the likelihood of a single taxonomy being established, yet the benefits of such a system are still valid as organizations increasingly collaborate for cyber defense (Harel et al. 2017). Based on our initial experience with a very large neural network, we designed a series of smaller neural networks that not only are faster but also provide better classifications for malware samples. Our experiments with a data set of more than 13,000 malware samples indicate the promise of our approach.

Further work to be done in the realm of using artificial intelligence (AI) in malware classification involves taking a deeper look into how individual vendors classify malware and using those features as the inputs to a neural network. Additionally, expanding the dataset of malware samples would allow the network to see more training data, producing a more accurate network.

## References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X. (2016) "TensorFlow: A system for large-scale machine learning," Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI '16, November, pp. 265 – 283.
- Cannell, J. (2013) "Tools of the Trade: Exploit Kits," Malwarebytes Labs (blog), February  
<https://blog.malwarebytes.com/cybercrime/2013/02/tools-of-the-trade-exploit-kits/>

---

<sup>4</sup> <https://www.tensorflow.org/>

<sup>5</sup> <https://caffe.berkeleyvision.org/>

- Claubaugh, Caroline, Dave Myszewski, and Jimmy Pang. (2020) "Neural Networks" Sophomore College, 2000, <https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/History/history1.html>. Accessed 22 Nov 2020.
- The Council of Economic Advisors (2018) "The Cost of Malicious Cyber Activity to the U.S. Economy", March. <https://www.whitehouse.gov/wp-content/uploads/2018/03/The-Cost-of-Malicious-Cyber-Activity-to-the-U.S.-Economy.pdf>
- de la Cuadra, F. (2007) "The genealogy of malware," Network Security, 2007 Volume 4, pp 17 – 20, April.
- Dumitras, T. and Neamtiu, I. (2011) "Experimental Challenges in Cyber Security: A Story of Provenance and Lineage for Malware," Proceedings of the 4th Workshop on Cyber Security Experimentation and Test, CSET '11, August.
- Harel, Y. Gal, I. B., and Elovici, Y. (2017) "Cyber Security and the Role of Intelligent Systems in Addressing its Challenges," ACM Transactions on Intelligence Systems and Technology, Vol. 8, No. 4, pp. 49:1 – 49:12, April. doi: 10.1145/3057729
- Hansen, S. S., Larsen, T. M. T., Stevanovic, M., and Pedersen, J. M. (2016) "An approach for detection and family classification of malware based on behavioral analysis," 2016 International Conference on Computing, Networking and Communications (ICNC), 15-18 February 2016
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. (2014) "Caffe: Convolution Architecture for Fast Feature Embedding," Proceedings of the 22nd ACM International Conference on Multimedia, MM '14, pp. 675 – 678, November.
- Landwehr, Carl E., Bull, Alan R., McDermott, John P. and Choi, William S. (1994) "A Taxonomy of Computer Program Security Flaws," ACM Computing Surveys, Vol. 26, No. 3, September, pp. 211 – 254.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998) "Gradient-based learning applied to document recognition." Proceedings of the IEEE, 86(11):2278-2324, November 1998.
- MalwareBytes Labs. (2020) "Trojan.VilseI", <https://blog.malwarebytes.com/detections/trojan-vilseI/>. Accessed 27 November.
- McCulloch, W., and Pitts, W. (1943) "A Logical Calculus of Ideas Immanent in Nervous Activity," Bulletin of Mathematical Biophysics, Vol. 5, No. 4, December, pp. 115 – 133. doi: 10.1007/BF02478259
- Reddemann, K. (2015) "Evolving Neural Networks in NPCs in Video Games" UMM CSci Senior Seminar Conference, May 2015.
- Marcos Sebastián, Richard Rivera, Platon Kotzias, and Caballero, J. (2016) "AVclass: A Tool for Massive Malware Labeling," Proceedings of the 19th International Symposium on Research in Attacks, Intrusions, and Defenses, RAID 2016, Springer, Cham, pp. 230 – 253, September 2016.
- Thanh, C. and Zelinka, I. (2019) "A Survey on Artificial Intelligence in Malware as Next-Generation Threats," Mendel Soft Computing Journal, Vol 25, No. 2, December, pp. 27 – 34.
- SonicWall. (2020) "Mid Year update 2020: SonicWall Cyber Threat Report", 2020, <https://www.sonicwall.com/resources/white-papers/mid-year-update-2020-sonicwall-cyber-threat-report/>
- Szor, P. (1998) "Attacks on Win32", Virus Bulletin Conference, October.
- Szor, P. (2005) The Art of Computer Virus Research and Defense. Addison-Wesley Professional.
- Wei, F., Li, Y., Roy, S., Ou, X., and Zhou, W. (2017) "Deep Ground Truth Analysis of Current Android Malware," Proceedings of the 14th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, DIMVA '17, July, pp. 252 – 276

# Employing Machine Learning Paradigms for Detecting DNS Tunnelling

Jitesh Miglani and Christina Thorpe

Technological University Dublin, Blanchardstown, Ireland

[jmig1995@gmail.com](mailto:jmig1995@gmail.com)

[Christina.thorpe@tudublin.ie](mailto:Christina.thorpe@tudublin.ie)

DOI: 10.34190/EWS.21.052

**Abstract:** Domain Name System (DNS) is an integral protocol which makes the resources available on the world wide web accessible. In recent times, there has been significant attention given to the development of different attack vectors against this protocol, out of which, DNS tunnelling is one of the most lethal attacks. Hackers use DNS tunnelling as a covert channel to cover their traces, exfiltrate data, and bypass the security policies enforced by the firewalls. Intrusion detection systems do not generally scan the DNS queries because it produces a lot of data which is not feasible to be analysed with a tool. This research proposes an active approach of detecting DNS tunnelling by capturing all packets in a local network and employing Machine Learning (ML) models to detect tunnelled data. The main aim of this research is to employ ML techniques to classify and separate DNS tunnelling packets from legitimate packets, compare the results generated based on their precision, detection time and computational effort, and determine an ideal solution for the problem. The research uses Ensemble Learning techniques which is a subset of Supervised ML to prepare a classifier that can detect DNS tunnelling. We also focus on the inherited implications of using ML which can be unknowingly faced while employing ML techniques, such as model overfitting, data bias, and complex and unrealistic model creation. The datasets used for this research are created using an emulated real-time network scenario. Our results show that ML classifiers can solve the problem of DNS tunnelling detection. Ensemble learning techniques outperforms decision tree by a large margin. Moreover, boosting models always produced better overall accuracy than bagging models, at the expense of a longer training time.

**Keywords:** DNS tunnelling, machine learning, attack detection

---

## 1. Introduction

Domain Name System (DNS) is a naming system which associates the domain names on the Internet to their actual addresses (Mockapetris, 1988). It basically translates the easy to remember alphabetic domain names to their numerical Internet Protocol (IP) addresses. The analogy of a phone book is often used to explain the purpose of DNS in the IP suite.

DNS uses a hierarchal and decentralized approach to resolve and obtain the actual address of a resource and thus making it more fault-tolerant and efficient. It was reported that more than 342.4 million domain names in total were registered by the end of 2018 (verisign 2018). So, having such a large amount of data at a single point can make it more prone to risks and errors. Another aspect of this story is the use of DNS. Each time a resource is accessed on the internet, a DNS query is generated against it to locate its actual address. This dependency on a crucial protocol has attracted hackers and threat actors since its origin.

DNS tunnelling is one such method in which the hacker creates a covert channel of communication. This communication channel can be used for various purposes ranging from data exfiltration to a Command and Control (CnC) centre in a botnet. The fundamental reason for establishing the communication channel over DNS is that Dynamic Host Configuration Protocol (DHCP) and DNS are the only two established and well-trusted protocols which are not monitored by firewalls (Ahuja, 2018). Another reason that aids in DNS tunnelling being undetected is in its volume. DNS queries are produced at such a high frequency that it results in a very large amount of data, making it unfeasible for any Intrusion Detection System (IDS) to analyse (Ahuja 2018). Therefore, detecting DNS tunnelling is the prime motivation behind this research.

The main objective, at a high level, that this research aims to resolve, is to classify DNS traffic based on whether it is coming from a DNS tunnel or a legitimate source. So, it can be said the Machine Learning (ML) technique needs to classify data into two different streams. This classification approach falls under supervised ML, a technique where *Input* and *Output* data are labelled. Through that labelled dataset, it deduces the basis of classification by identifying patterns for future data processing (Caruana 2006). Although, the concept of supervised learning is comparatively straight forward, it provides a variety of ML models for classification. For this research, a sub-branch of the supervised learning technique is considered for DNS tunnelling detection,

namely Ensemble Learning (EL). The EL model is an advanced version of the primitive supervised learning technique. It combines various weak classifiers, offered by supervised learning, together to form a strong classifier. EL offers a couple of different approaches as well to create a strong classifier. In this research, all those approaches will be discussed, used, and compared to find a robust solution for DNS tunnelling detection.

Currently, DNS tunnelling is one of the few attack vectors for which there are no designated and efficient detection methods. DNS is required to access anything on the Internet, and hence, the DNS port (port 53) cannot be blocked by the firewall. Furthermore, the IDS cannot analyse data generated by DNS queries because of its volume. One of the prime reasons for DNS tunnelling being undetected is that DNS queries generate datasets large enough that cannot be processed by a normal IDS. ML driven approaches can help to ensure that anomalies are detected even if they are obscured by a large volume.

The aim of this research is to first develop an architecture and a dataset specifically for DNS tunnelling which could be used for further research in this area. Then, applying different ML algorithms using that dataset to detect DNS tunnelling and compare the efficiency and accuracy of the algorithms based on their outputs to provide a reliable solution for this problem.

## **2. Background and related work**

Ensemble can be defined as viewing a group of items as a whole rather than treating them individually. Ensemble Learning (EL) algorithms are based on this definition, in the sense that a group of weak learners is viewed and clubbed together to form a strong learner, thus improving the overall accuracy of a machine learning model (Geron 2019). That is, EL aggregates the predictions derived from various predictors of a classifier to get better prediction than from an individual predictor and removes most of the machine learning error paradoxes of noise, bias, and variance. EL has two common categories, namely Bagging and Boosting. Bagging is an ensemble learning algorithm in which the predictor or the classifier remains the same in each iteration but is sampled on different subsets of the training set. Once all classifiers are trained, EL then makes new predictions on testing dataset by combining all the predictions using an aggregation function, which is the reason why bagging is also known as Bootstrap Aggregation. Boosting or Hypothesis Boosting is similar to bagging in that it combines the predictions of a few weak classifiers to prepare a strong classifier. However, one major difference is that with Bagging all classifiers are trained parallelly and at the end the results are aggregated, whereas Boosting is a sequential process that takes the output from one weak classifier to train the next one.

There has been a lot of research conducted towards the collaboration of ML models and cybersecurity to eliminate the Big Data problems in the last few years. One of the first contributions in this domain was proposed by (Dusi 2008). They tried to detect the presence of encrypted tunnels in each network by using a rule-based classification model to separate legitimate traffic from tunnelled traffic. The researchers used their previous work as a basis, in which they classified benign traffic using statistical fingerprinting into their respective application layer protocols (Crotti 2007). Based on the promising results computed in (Dusi 2008), they extended their research and introduced a core ML technique for detecting on encrypted tunnels (Dusi 2009). The research was based on the same premise, but this time only packet size and inter-arrival time of the packets were considered to compute statistical fingerprints. The scope of the research was increased as it used two different tunnelling mechanisms, HTTP and SSH to detect legitimate traffic.

(Aiello 2013, 2015, 2016) proposed to detect DNS tunnelling using a supervised learning algorithm in ML. The researchers propose that data tunnelled by DNS protocol can be detected by using the statistical features of the contents in each DNS query and answer pair and feeding them into a supervised learning model which in this case would be Bayes Classifier. (Buczak 2016) worked towards the same direction of detecting DNS tunnelled data but using a different ML model. They used the Random Forest model to detect tunnelled data. (Bubnov 2018) recently conducted similar research to (Buczak 2016) and used four different tunnelling tools to simulate the DNS tunnels but instead of using Random Forest, the researcher used Feedforward Neural Network technique, which is one of the latest ML techniques to detect the tunnelled data. (Engelstad 2017) pointed out the fact that DNS tunnelling is a problem that can be seen in both thick and thin clients. More recently, (Watkins 2017) discussed DNS tunnelling as a Big Data problem and tried to solve that using Semi-supervised ML techniques. A more recent contribution also includes the binary classification of DNS tunnelling (Liu 2017). The researchers used supervised learning classifiers like Decision Tree, SVM and Logistic regression to classify legitimate and tunnelled packets.

The efficiency and accuracy of any given ML model depend predominantly on the features extracted from the dataset. Considering the work done by various researchers towards the same direction, (Sammour 2018), published a paper which talks about the important features that can be used for DNS tunnelling detection and different analysis types for the same.

This research combines the knowledge of the state of the art to design a common DNS tunnelling architecture to generate datasets for the research in this area of concern. We use an ensemble learning technique, which to the best of our knowledge, has not been used in any of the existing literature. As the problem of DNS tunnelling is being tackled as a Big data problem, there are certain ML obstacles like model overfitting and data bias that are also inherited if the models are not trained properly. The state of the art presented does not account for that; to rectify that, this research proposes four different model training strategies, which could provide a solution for the ML problems of model overfitting and data bias also.

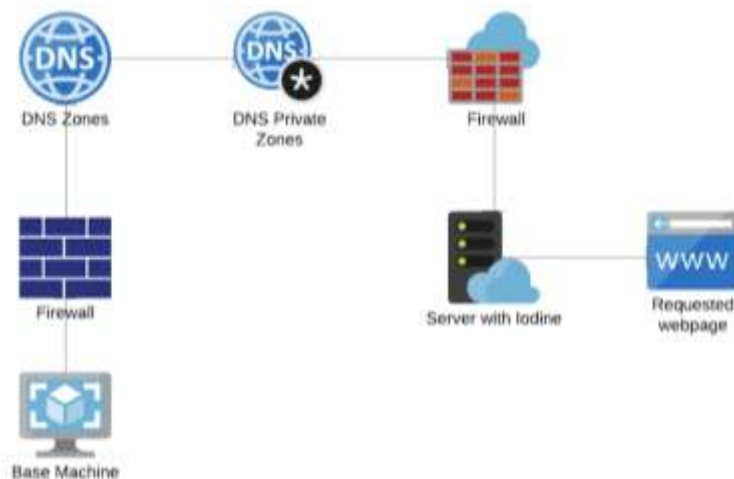
### 3. Methodology

The first phase of the research involved the experimental setup which was required to collect data for analysis. This involved emulating an active DNS tunnel to gather tunnelled packets. The tunnel was emulated over a variety of application layer protocols like HTTP, HTTPS, UDP and FTP to make the final dataset richer and more diverse. Parallely, packets for legitimate traffic were pulled from an online repository. The features extracted from those packets were based on the information correlated from the literature review. The set of features extracted where a combination of different features used by researchers while conducting their research on a similar matter. The output of this phase was a dataset of 100,000 records, which was pre-processed and labelled in preparation for the next phase.

The second and the last phase involved training a variety of ML classifiers (including Decision Tree, Random Forest, Gradient Boosting Algorithm, AdaBoost, Neural Networks and K-Means algorithm) on a portion of the dataset created and testing the performance of those models. This phase not only involved building ML models to classify the data, but it also involved the analysis of their results. The results were further used in this phase to compare all models among themselves and tune the models for a better classification. The results gathered were also compared using different factors like model's accuracy and precision and the computation effort required.

#### 3.1 DNS tunnelling architecture

One of the shortcomings of researching this area is the absence of a benchmarking dataset over which the research can be compared. Due to this reason, one part of this research was dedicated to the creation of a well-structured architecture that could be used to generate a dataset that could be used in the future for researching this domain (see Figure 1).



**Figure 1:** DNS tunnelling architecture

The DNS protocol, just like any other application layer protocol like HTTP, works on a query and response basis. The protocol sends out a query which just carries the name of the server and the response received brings the

numeric address mapped for that name on the network. However, using the same principles of this protocol the queries and response can be tweaked. In DNS tunnelling, instead of sending out a query with a name of the server, the attacker can send out GET command encapsulated in the DNS packet and receives a POST response encapsulated in the DNS response packet (instead of a numeric address).

The technical setup for this attack vector requires a separate domain name server to receive and respond to those encapsulated queries, a valid hostname that would be pointing towards the DNS server, and finally a tunnelling tool that would work in client-server model installed over the server. This tunnelling tool converts the encapsulated queries into meaningful data and then encapsulates the valid responses back into DNS response packets. Iodine was selected for the DNS tunnelling in this work (Erik Ekman 2015).

### **3.2 Data collection**

To diversify the dataset and make the ML model more robust in identifying tunnelled data, it was important to have a certain amount of width in the dataset. To achieve this, different application layer protocols were used over the tunnel, including HTTP, HTTPS, FTP and UDP. The DNS tunnel was used to access the web. Firstly, the HTTP protocol was tunnelled over DNS. A list of sites still using HTTP protocol was downloaded from [scratchpad.eu](http://scratchpads.eu) (<http://scratchpads.eu/explore/sites-list>). The list consisted of over 1000 URLs for different websites and their domains that still use the HTTP protocol. Using the DNS tunnel, these websites were accessed, data was entered to fields and some PDFs were downloaded in the process too. The tunnel was emulated until 30,000 DNS packets were captured for the HTTP protocol. A similar approach was used while tunnelling HTTPS protocol over DNS. While collecting HTTPS tunnelled data, there was less limitations as most websites that are regularly being used like Wikipedia, Google, Twitter use HTTPS. The collection duration and the target number of packets were the same. For the collection of UDP data tunnelled over DNS, the best option was to stream videos on different websites like YouTube and Dailymotion.com. The only problem faced during the collection of UDP tunnelled data was because of the low throughput of the tunnel, the videos played for streaming took a long time to load. Finally, FTP was tunnelled over DNS; this was achieved by using Proxychains. Files of different sizes were downloaded from a public test server created by <http://speedtest.net>. The duration of tunnelling FTP over DNS was equal to the duration of HTTP and HTTPS tunnels. The final count of the tunnelled dataset was 100,000 packets approximately. The data collected was in .pcap format, from which the features were extracted and the dataset was cleaned and transformed in .CSV format.

The collection of legitimate DNS queries and responses was equally important, and it had to be almost equal to the amount of tunnelled DNS queries and responses which means accessing approximately 50,000 URLs on the web. Luckily, to avoid the tedium, unlike the absence of a proper dataset for DNS tunnelling data, there were many .PCAP files available on the internet for legitimate DNS queries on various dataset repositories. One of which was Mendelej; this contained a dataset called 10 Days DNS Network Traffic from April-May 2016 created by (Singh 2016). The dataset had 10 days' worth of DNS traffic recorded from the edge routers of a college campus that had more than 4,000 active users each day. For this research, a packet capture of the first day was more than sufficient (Day 024042016, 2016).

### **3.3 Feature extraction**

Six features were selected from the Payload of a packet by combining the most common and important features used in the existing literature: DNS Query name, Entropy, DNS Query Length, DNS Response Length, IP Packet Size (Response), DNS Query Type, Interarrival Time of Query, and Response. The combination of these six features and their intended purpose in training the model is unique. The main reason for selecting each of these features is defined in table 1.

### **3.4 Data preparation and pre-processing**

Once all features were extracted, a separate column was created; in the case of legitimate packet records, that column was filled with 0 and for tunnelled packets the column was filled with 1. The prepared dataset had to be divided into training and validation datasets. Each model had to be trained in two separate ways. In the first approach, both training and validation datasets had a balanced ratio of legitimate and tunnelled entries. In this approach, the training dataset had 80% of the actual dataset and the validation dataset had the remaining 20% of the original dataset. In the second approach, the original dataset was divided into two equal parts. The first part had an equal number of legitimate and tunnelled records in it and was used for model training. The second

part was then further split into three separate parts in which tunnelled and legitimate queries were in the following ratios 1:10, 1:100, 1:1000. These three parts were then used for model testing in the second approach.

**Table 1:** Feature selection rationale

<b>Feature Name</b>	<b>Selection Rationale</b>
<i>DNS Query Name Entropy</i>	As in machine learning the DNS query name would be considered just as another string value, it would not have a suitable impact on the model training process. Whereas by calculating its entropy would make a numeric and valid feature out of it. The entropy would be a stronger feature as DNS tunnelled queries are Base32 encoded which make them more random than legitimate DNS queries hence there would always be a significant difference between entropy values of the two queries.
<i>DNS Query Length</i>	The Query length passed to the DNS would be significantly larger than regular queries for a tunnelled packet as an entire fragmented IP packet would be encoded inside it.
<i>DNS Response Length</i>	The Response length of the DNS response would also be larger than regular responses for a tunnelled packet as instead of an IP address for a resource on the network, the response section would have an entire fragmented IP packet would be encoded inside it.
<i>IP Packet Length</i>	There would be a huge difference of size of the entire packet size in between a tunnelled and a legitimate packet.
<i>DNS Query Type</i>	The packets tunnelled over DNS cannot use regular DNS records. They can either use TXT or NULL record types [60].
<i>Interarrival Time</i>	The interarrival time between query and response would vary as the latency between the two would increase in case of DNS tunnelling.

### 3.5 Model training

For this research, four models were selected: one from the existing literature, Decision Tree; two from Ensemble learning, Gradient Boosting Algorithm and Adaptive Boosting Algorithm; and one which was common in both, Random Forest algorithm.

In the first model training strategy, the four models were trained and tested over the balanced dataset. The entire dataset was loaded and was divided into two parts one for training and one for testing. Once the dataset was divided, functions for the four models Decision Tree, Random Forest, Gradient Boosting Algorithm and AdaBoost were called from the Sci-Kit Learn library. No hyperparameter tuning was done for any of the models. Then using K-folds Cross validation (K=5), the models were evaluated, and their accuracy was recorded.

In the second strategy, training used two separate methods. In the first method, Gaussian noise (to ensure realistic randomness) was added to the entire dataset and then the dataset was divided using train test split function into 20% of validation dataset and 80% of training dataset and then the remaining steps in the first strategy were followed to evaluate the model performance. In the second approach, the datasets prepared in the second method of data splitting were used. The noise was only added to the training dataset and the testing dataset was held back. The models were trained over noisy data and tested over regular data and evaluated using K-folds cross validation (K=5).

In the third strategy, data normalization was used for feature scaling as there was a large difference in the values among various features for model optimization. Such huge differences can create a sense of bias in the models and can cause the model to memorize instead of learning. This huge difference between features and classes was removed using data normalization. The features were normalized between 0-1, the difference between the classes and features was minimized and the relationship between features was still intact. The proceeding model training and validation approach used was like that of the second strategy.

The second and third strategies were employed to inspect whether the data collected was faulty in any manner. However, the results did not waiver significantly, thus alleviating the suspicions of having a faulty dataset. This led to the fourth strategy, in which the models were examined for overfitting and creation of unrealistic complex models. Previously, no hyperparameter tuning was done for any models. In this strategy, an optimal value for



hyperparameters for each model was derived and the models were properly tuned. To avoid overfitting and computational inefficiency, each model was tuned separately by looking at the performance of each hyperparameter during model training and testing by plotting their graphs. The rationale behind selecting the optimal value was that the accuracy should be maximum given there is a minimum difference between training and testing accuracy.

#### 4. Analysis

This section discusses the parameters over which the performance of each classifier were evaluated and compared. The confusion matrix provides the required values to calculate other parameters like precision, recall, f-measure and accuracy. Based on the derived values different plots such as PR curves and ROC curves are plotted.

This model training strategy was implemented to restrict models from creating unrealistic and complex models at the cost of high net accuracy. With this, the optimal values for each models' hyperparameters were found and set before model training and testing, and the architecture of the models were fixed. The training and testing of each model were done similarly: first, the models were trained and tested over a balanced dataset where 80% of the dataset was used for model training and the remaining 20% was treated as the validation dataset, and net accuracy for each model was calculated. After this, the models were trained and tested over three sets of imbalanced datasets to find each models precision, recall and f-measure values.

##### 4.1 Net accuracy

**Table 2:** Net accuracy - model training strategy #4

Classifier	Training Time	Training Accuracy	Testing Accuracy (K-folds CV)
Decision Tree	0.023871	0.964307	0.964307
Random Forest	0.117697	0.991617	0.985122
Gradient Boosting	0.296852	0.997147	0.997731
AdaBoost	0.336374	0.998057	0.998071

Because the hyperparameters were initially tuned, the training time of all models drastically dropped. The training and testing accuracy were quite similar to each other but had dropped a little for each model. However, one thing was constant in this model training strategy as well; Decision Tree had the lowest training and testing accuracy among all models. Both boosting algorithms had the highest accuracies in both categories (it was in the first model training strategy), which was then followed by the Random Forest classifier.

##### 4.2 Confusion matrix

The confusion matrix for Decision Tree had the greatest number of false negatives which were 602 in total. In total, Decision Tree had 639 misclassified records from the validation dataset, giving it an overall accuracy of 96.4%, the lowest among all for models. This was followed by the bagging algorithm, Random Forest. It had 69 false negatives and 66 false positives and had an overall accuracy of 98.5%. The two boosting algorithms, yet again, had a similar performance. Gradient boosting outperformed AdaBoost this time as it had 24 misclassified records whereas AdaBoost had 28 misclassified records which made their overall accuracy of 99.86% and 99.84%, respectively. All four models performed quite well based on their overall accuracy, but Decision Tree and Random Forest had the highest numbers of false negatives in their classification report. This seemed a trade-off, as in this strategy they had their lowest recorded training time as compared to all other model training strategies.

##### 4.3 Precision, Recall and F-measure

The Precision and Recall were between 0.995-0.998 and 0.934-0.994, respectively (in all classifications when the dataset was balanced). The precision behaved the same as the last strategies were there was a drop in the recall value because of the increase in the number of false negatives for each model. This changed when the classifiers were tested over three different sets of imbalanced classes.

**Table 3:** Precision, Recall & F-measure – training strategy 1

Classifier Legitimate = 0; Tunnelled = 1	Precision	Recall	F-measure
Decision Tree: 0 Decision Tree:1	0.9931 0.99679	0.9997 0.9305	0.99639 0.9625
Random Forest:0 Random Forest:1	0.9999 0.96662	0.99655 0.999	0.99822 0.98254
GBA:0 GBA:1	1 0.99354	0.99935 1	0.99967 0.99676
AdaBoost:0 AdaBoost:1	0.99995 0.999	0.9999 0.9995	0.99992 0.99925

**Table 4:** Precision, Recall & F-measure - training strategy 2

Classifier Legitimate = 0; Tunnelled = 1	Precision	Recall	F-measure
Decision Tree: 0 Decision Tree:1	0.99955 0.96954	0.9997 0.955	0.99963 0.96222
Random Forest:0 Random Forest:1	0.9999 0.95652	0.99955 0.99	0.99972 0.97297
GBA:0 GBA:1	1 0.95694	0.99955 1	0.99977 0.978
AdaBoost:0 AdaBoost:1	0.99995 0.99005	0.9999 0.995	0.99992 0.99252
Classifier Legitimate = 0; Tunnelled = 1	Precision	Recall	F-measure
Decision Tree: 0 Decision Tree:1	0.9998 0.72727	0.9997 0.8	0.99975 0.7619
Random Forest:0 Random Forest:1	1 0.83333	0.9998 1	0.9999 0.90909
GBA:0 GBA:1	1 0.95238	0.99995 1	0.99997 0.97561
AdaBoost:0 AdaBoost:1	1 0.95238	0.99995 1	0.99997 0.97561

## 5. Results

The four models that were selected for this research were Decision Tree, Random Forest, Gradient Boosting Algorithm and AdaBoost. Out of these four models, all except Decision Tree were Ensemble learning techniques. Based on our findings, it can be said that out of the four models, Decision tree had the lowest performance over all the metrics. Ensemble learning is further divided into two categories: Bagging and Boosting. Random Forest follows the bagging technique, whereas the remaining two models (as their name suggest), fall under boosting technique. If the results of all the performance metrics of all four datasets are considered, the Boosting technique outperformed the Bagging technique in DNS tunnelling detection. The two boosting models Gradient Boosting and AdaBoost had a very similar performance in all model training strategies. However, the training time for Gradient boosting was higher than AdaBoost and the precision, re- call and f-measure values were higher for AdaBoost in imbalanced datasets. This means AdaBoost outperformed Gradient Boosting Algorithm.

## 6. Conclusions and future work

The research was conducted with one main goal of classifying and separating out tunnelled packets from legitimate packets in each dataset. To achieve this goal, the entire project was divided into two phases. The first phase primarily involved the creation of a dataset. The first step in doing so was to establish a stable DNS tunnel which was done using public cloud infrastructure and a DNS tunnelling tool called Iodine. Once the tunnel was configured and was stable enough for use, different application layer protocols were tunnelled using that. Over 100,000 DNS queries and responses were captured, and an equal number of legitimate DNS packets were also collected from Mendeley (Singh 2018). Based on the inputs of phase one, a list of six key features were created and those six features were extracted from both tunnelled and legitimate query and response pairs. This was the first dataset that was created and had a 50% mix of tunnelled and legitimate queries. After this, three more datasets were created but this time the ratio of tunnelled and legitimate queries was not balanced. The three imbalanced datasets had a ratio of 1:10, 1:100, 1:1000 tunnelled to legitimate queries. The reason behind tunnelling application layer protocols and creating three imbalanced datasets was to prepare a model that is robust enough, that it could perform well in real-time scenarios. These datasets were then used to train various models in the second phase. Based on the analysis of the second phase it can be said that the ML classifiers can solve the problem of DNS tunnelling detection. As far as the best classifier is considered, ensemble learning technique outperforms Decision tree by a large margin. Moreover, between bagging and boosting learning models, boosting models always had better overall accuracy, at the expense of a longer training time. For future work, the dataset can be increased and improved using the same architecture and varying tunnel parameters.

## References

- Aiello, M., Mongelli, M. and Papaleo, G. (2013), July. Basic classifiers for DNS tunneling detection. In *2013 IEEE Symposium on Computers and Communications (ISCC)* (pp. 000880-000885). IEEE.
- Aiello, M., Mongelli, M. and Papaleo, G. (2015). DNS tunneling detection through statistical fingerprints of protocol messages and machine learning. *International Journal of Communication Systems*, 28(14), pp.1987-2002.
- Aiello, M., Mongelli, M., Cambiaso, E. and Papaleo, G. (2016). Profiling DNS tunneling attacks with PCA and mutual information. *Logic Journal of the IGPL*, 24(6), pp.957-970.
- Buczak, A.L., Hanke, P.A., Cancro, G.J., Toma, M.K., Watkins, L.A. and Chavis, J.S. (2016), April. Detection of tunnels in PCAP data by random forests. In *Proceedings of the 11th Annual Cyber and Information Security Research Conference* (pp. 1-4).
- Caruana, R. and Niculescu-Mizil, A. (2006), June. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161-168).
- Anjum Ahuja (2018). Plight at the end of the tunnel; Aug 14 2018. URL <https://www.endgame.com/blog/technical-blog/plight-end-tunnel>.
- Erik Ekman (2015). Iodine; 2015. URL <https://github.com/yarrick/iodine>.
- Engelstad, P., Feng, B. and van Do, T., (2017), March. Detection of DNS tunneling in mobile networks using machine learning. In *International Conference on Information Science and Applications* (pp. 221-230). Springer, Singapore.
- Dusi, M., Crotti, M., Gringoli, F. and Salgarelli, L. (2008), May. Detection of encrypted tunnels across network boundaries. In *2008 IEEE International Conference on Communications* (pp. 1738-1744). IEEE.
- Crotti, M., Dusi, M., Gringoli, F. and Salgarelli, L., (2007). Traffic classification through simple statistical fingerprinting. *ACM SIGCOMM Computer Communication Review*, 37(1), pp.5-16.
- Dusi, M., Crotti, M., Gringoli, F. and Salgarelli, L., (2009). Tunnel hunter: Detecting application-layer tunnels with statistical fingerprinting. *Computer Networks*, 53(1), pp.81-97.
- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.
- Liu, J., Li, S., Zhang, Y., Xiao, J., Chang, P. and Peng, C., 2017, August. Detecting DNS tunnel through binary-classification based on behavior features. In *2017 IEEE Trustcom/BigDataSE/ICSS* (pp. 339-346). IEEE
- Manmeet Singh (2018). 10 days dns network trac from april-may, 2016, 2018. URL <https://search.datacite.org/works/10.17632/ZH3WNDDZXY>
- Manmeet Singh. (Day 024042016, 2018). URL <https://search.datacite.org/works/10.17632/ZH3WNDDZXY>.
- Mockapetris, P. and Dunlap, K.J., (1988), August. Development of the domain name system. In *Symposium proceedings on Communications architectures and protocols* (pp. 123-133).
- Sammour, M., Hussin, B., Othman, M.F.I., Doheir, M., AlShaikhdeeb, B. and Talib, M.S. (2018). DNS Tunneling: a Review on Features. *International Journal of Engineering and Technology*, 7(3.20), pp.1-5.
- Verisign (2018). The verisign domain name industry brief. Technical report, Sept 30 2018. URL [https://www.verisign.com/en\\_US/domain-names/dnib/index.xhtml](https://www.verisign.com/en_US/domain-names/dnib/index.xhtml).
- Watkins, L., Beck, S., Zook, J., Buczak, A., Chavis, J., Robinson, W.H., Morales, J.A. and Mishra, S., 2017, January. Using semi-supervised machine learning to address the big data problem in DNS networks. In *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 1-6). IEEE.
- Yakov Bubnov (2018). Dns tunneling detection using feedforward neural network. *European Journal of Engineering Research and Science*, 3(11):16–19, Nov 21, 2018. doi: 10.24018/ejers.2018.3.11.963.

# Analysis of API Driven Application to Detect Smishing Attacks

Pranav Phadke and Christina Thorpe

Technological University Dublin, Blanchardstown, Ireland

[phadke.pranav09@gmail.com](mailto:phadke.pranav09@gmail.com)

[christina.thorpe@tudublin.ie](mailto:christina.thorpe@tudublin.ie)

DOI: 10.34190/EWS.21.051

**Abstract:** In the past decade, the use of mobile smart phones has increased exponentially. The pervasiveness of these devices has motivated criminals to design ways to exploit the mobile technology to obtain confidential information or to execute malicious software. The term used to describe these social engineering attacks using mobile phone technology is Smishing, which is a play on the previously well-known Phishing attack perpetrated over email. Smishing uses Short Message Service or SMS as its attack vector to send malicious Uniform Resource Locators (URLs) along with the text message. Users are more aware of phishing emails and there are a lot of detection mechanisms developed to avoid such attacks. However, SMS is often neglected, and considering the small size of the mobile screen compared to a computer, it is difficult to detect and manually verify a phishing URL which is sent in a text message. When clicked, a smishing URL can either redirect the user to some phishing page or try to install a malicious payload on the mobile phone. Both scenarios are risky and can cause potential loss. The aim of this research is to develop a new application to detect Smishing attacks on Android devices by integrating existing phishing Application Program Interfaces (APIs) in a prototype application. The application is designed to run in the background and verify whether the URL in the text message is phishing or not. Five freely available APIs were tested on a dataset of 1500 URL to compare them in terms of accuracy and latency. Our results show that the VirusTotal API gives the most accurate detection rate of 99.27%, but the slowest response time of 12-15 seconds per query; the Safe-Browsing API, gives an 87% accuracy with 0.15ms response time. For time sensitive applications, Safe-Browsing would provide the best solution, however, for security sensitive applications, Virus Total would be a better option.

**Keywords:** phishing, smishing, application programming interface, cybersecurity, attack detection

---

## 1. Introduction

Phishing is a social-engineering cybercrime in which the attacker uses email, telephone, or a text message as the attack vector to send seemingly harmless links which redirect the users to a malicious site. The aim of the spoofed sites is to lure victims into providing sensitive and confidential information or to prompt them to download malicious software (Dhamija, Tygar, and Hearst, 2006) (Jagatic et al, 2007). This can have devastating results on both individuals and organizations. For an individual, the attacker can use their personal information to seek monetary benefit, obtain ransom, etc. Whereas for organisations, it could be used to defame the organisation. If confidential data is compromised, it could cost the organisation in various ways such as heavy financial loss, reputation, or a loss of share value and customer confidence.

Mobile phones have now become a necessity rather than a luxury; it is estimated that more than 5 billion people have mobile phones (Taylor and Silver 2019). Mobiles have evolved significantly from just being a telephony device, to now being a powerful survival tool. With the current technology, and ease of access, it is the first choice amongst gadgets for everyone, encompassing both younger and older generations.

However, due to its popularity amongst the masses, and our growing dependence on mobile phones, they have also become an easy target for cyber criminals to carry out various types of attacks. People use phones to store personal information (credentials for various apps, bank details, etc), hence, it is a massive opportunity for hackers to target and exploit this interface and data. Many users synchronise their office and personal emails on mobile devices and access it without entering a password, some users store bank credentials and personal information on notepads, and finally, the hard disk may contain personal data such as photos and videos, text messages, calendar information, etc. All this information is critical and confidential and if compromised, can be dangerous for the user or organization (Pieterse, Olivier, and Heerden 2019).

Smishing is a type of social-engineering attack where the attacker uses SMS as an attack vector to send a phishing link which would direct the user to some malicious site or execute a payload on the device. Our research focuses on smishing and explores a mitigation technique using phishing detection APIs. We built a mobile application for the Android platform to validate our proposal and answer the following research questions: (1) Can a phishing detection APIs be used to detect smishing? (2) Which phishing detection APIs are the most accurate and suitable for this prototype? (3) Can an API be integrated in the background process/broadcast receiver?

The aim of this research is to develop a new application to detect Smishing attacks on Android devices by integrating existing phishing Application Program Interfaces (APIs) in a prototype application. The application is designed to run in the background and verify whether the URL in the text message is smishing or not. Five freely available APIs were tested on a dataset of 1500 (malicious and legitimate) URLs to compare them in terms of accuracy and latency. Our results show that the various APIs perform differently in terms of accuracy and latency, where the most accurate solution is the slowest solution. This work may be a valuable resource for developers who need to select the most suitable API for their security and performance requirements.

## 2. Background and related work

An API provides a method to integrate an application with an external service, connect a particular interface to an application and share the data with customers or external users. This improves application implementation whilst keeping the time and monetary constraints within limit. API research has been a very active area over the past decade (Ofoeda, Boateng, and Effah, 2019). Benefits of developing with APIs include: (1) Any organization can customize the response according to their needs and share it with the end users. (2) Applications change over the time and API offers swift data migration and adaptation. (3) APIs offer a wide range of efficiency and quick response while performing dedicated tasks, making them efficient.

With the growing use of APIs in enterprise level applications, there were many APIs developed in the cyber-security domain e.g., phishing detection APIs. So, the objective of this research is to test various, easily available, and free APIs which would validate the given URL to classify it as either a phishing or legitimate URL and later, integrate any one of the available API in the prototype application which would successfully detect smishing attacks.

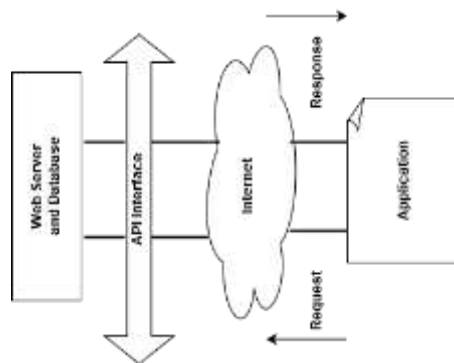


Figure 1: API lifecycle

Figure 1 depicts the general API lifecycle. An interface, e.g., a mobile application that wants to request an authorization token, would send an API request to the API interface through the internet. When the API request reaches the interface, the API web server generates the token and sends it to the client application in the form of a response. This is the general workflow mechanism of an API. The five APIs identified for this research are detailed in the following subsections:

### 2.1 Safe Browsing API

Google has created its own API called Safe-Browsing to identify phishing URLs. Google constantly updates their list of unsafe and malicious web resources like phishing and malicious sites, blacklisted domains, and sites that host unwanted software. By implementing Google's safe browsing API, one can let the client application check the URL against Google's list of unsafe web services. Social Engineering sites like phishing and deceptive sites can be checked against the list which is updated frequently (developers.google.com, 2019)

### 2.2 Phishtank API

PhishTank, operated by OpenDNS, is a community driven website for submitting, verifying, sharing, and tracking phishing URLs. PhishTank gives precise, noteworthy data to anybody attempting to distinguish between a phishing and legitimate URL. Anyone can submit a suspected URL on the PhishTank website. Registered users cast their votes against the submitted URL, which determines if the URL is classified as a "Phish" or "No Phish". This lessens the probability of having false negatives and improves the general aggregation of phishing

information. Votes are calculated based on duration, latency, and accuracy in relation to other messages to prevent scammers and maintain the integrity of the voting community (phishtank, 2019) (spamfighter, 2019).

### **2.3 UrlScan.io**

Urlscan.io is an interface used to browse and analyse a particular URL. The core methodology behind developing this interface was to give amateur users an insight into how a particular URL or website functions, the functions called, and requests sent in the background. When the users submit a URL to the web interface or to the API, an AI process browses the URL in the same way a user would browse. During the browsing session the automated mechanism gathers information about the website, page by page, as it traverses throughout the website. This information includes domain name, various IP addresses involved, and client-side resources requested like jQuery files. The service will also take a snapshot of each page as it traverses through. Upon completion, urlscan.io takes assistance of external resources for further verification of whether the page is malicious or legitimate. By using CryptoJacking, it also checks if the URL is an attack to mine cryptocurrencies (urlscan.io, 2019).

### **2.4 CyberFish**

Cyber Fish is an anti-phishing mechanism which integrates Artificial intelligence and computer vision. Computer Vision enables the computers to see, identify, and process a particular image just like human beings (Shi and Li 2018). Artificial intelligence helps in enhancing the performance through robust acceleration. This helps develop improved predictive modelling to identify the attacks. With this combination, it helps CyberFish to detect signature-less attacks (cyberfish, 2019).

### **2.5 VirusTotal**

VirusTotal is a free service to verify and analyse URLs, hashes, and malicious files. It integrates approximately 70 different antivirus scanners and services to identify and detect trojans, viruses, or worms in a file or URL. Along with the anti-virus integration it also uses web system scanners to detect any vulnerabilities and identify the malicious code in the file (virustotal.com, 2019).

Using existing APIs ensures that the detection process is efficient. The APIs are also updated quite often, which ensures that the prototype can defend against even the latest attack projected by the attacker. Furthermore, these specific five APIs selected are widely used and free, and do not have any hidden charges/membership requirements. This ensures that the product developed from this methodology is easily available for integration, helping to reduce the cost involved in developing a full-fledged product.

Mobile Phishing Attacks and Mitigation Techniques by Hossain Shahriar, Tulin Klintic, Victor Clincy (Shahriar, Klintic and Clincy 2015) discusses how mobile devices are the primary focus of attackers and how mobile phishing has emerged as a threat in today's modern world. The paper notes that small size of the mobile phone screen, which in turn could result in URLs displaying only partially, are a reason for widespread mobile phishing attacks. Cybercriminals rely on the small size of the screen, as this makes it harder for the user to check the validity of a URL sent via SMS. Moreover, short URLs are now being used by hackers to hide the true identity of the URL. The paper also focuses on how the fake URLs are created to imitate the look of legitimate URLs making it harder for a layman to detect phishing scams.

(Longfei Wu, Xiaojiang Du, Jie Wu, 2014) propose a novel lightweight anti-phishing application developed on an Android operating system called MobiFish, which checks the credibility of applications and web pages by evaluating them with the actual legitimate applications and web pages. MobiFish tries to solve the phishing problem without being dependent on the heuristic approach or any machine learning technique. MobiFish uses optical character recognition (ORC) technique to extract text from screenshots of a web interface along with the URL. These texts are matched to the one that users are redirected to, and if there is an anomaly then the current page is regarded as a phishing page.

Since SMS attacks on mobile screens are a preferred mode of attack, aided by the small size of the screen, we aim to create a prototype smishing attack detector which eliminates the threat at the URL level itself. Our approach will ensure that malicious or smishing URLs are not opened by the end user, thereby preserving their integrity. In addition to this, using APIs as the detection mechanism offers benefits due to their versatile nature.

APIs can offer multiple detection mechanisms like machine learning, blacklisting, heuristic, or artificial intelligence thus offering more flexibility in the detection approach rather than depending on one particular approach as seen in the previous research.

### 3. Prototype overview and system implementation

The proposed prototype was designed to detect smishing attacks which are particularly focused on mobile phones. The design is based on a broadcast receiver which runs constantly in the background and listens to incoming messages on the device. A broadcast receiver is an android system component which listens to any system event, in this case, an incoming SMS (developer.android.com, 2019). The receiver listens to incoming messages and extracts the URL from the plain text. APIs were used for URL verification since they offer multiple pathways of detection (AI, Backlisting, and Machine Learning) and the URL payload is not executed on the local machine or mobile device. In this interface there is no need to open the URL on the browser to invoke any kind of verification.

The broadcast receiver is developed in such a way that it would keep running in the background and fetch any inbound SMS. Upon fetching the message and extracting the URL it would then forward it to the API for verification. This optimizes the response time for verification and nullifies the chances of any sort of malware payload, if present, getting executed on the mobile device. Moreover, this application consumes less memory and battery life since it does not run in the foreground all the time (Dawn Griffiths, 2017). Figure 2 depicts the flow of the prototype and functions of each module.

#### 3.1 Message reader module

This module is programmed to listen to incoming messages and read it in the background. A smishing message would normally be a combination of text and URL. The program reads the message as a whole string including the URL. So, it was important to extract and split the URL contents from the string.

#### 3.2 API forwarder module

This module runs in an asynchronous mode and takes the URL as an input parameter. Along with the URL, the API key, response type, and authorization headers are all added in the request object and sent to the API for verification. The PhishTank API was integrated for URL verification in the prototype implementation.

#### 3.3 Alert notification module

This is the final module of a smishing based prototype engine. The API processes the URL and sends a response object back to the system. The response normally has the data related to the URL (date, timestamp, valid phish or not, location, etc). If the URL is verified by the API as a valid phish then an alert notification is sent across the screen to the end user. This notification warns the user about a potential malicious threat present in the SMS that prevents the user from opening it.

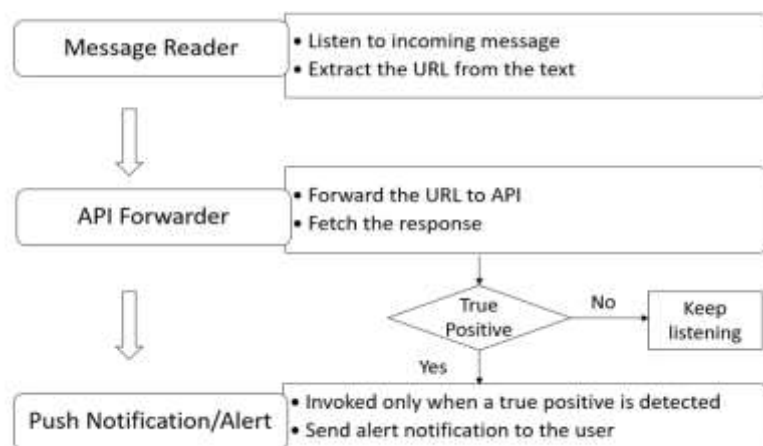


Figure 2: App flowchart

## 4. Methodology

A dataset comprised of 1,500 URLs was used to test the accuracy and performance of the five APIs. It contains randomly selected 900 phishing URLs taken from the OpenPhish feed, and 600 legitimate URLs taken from University of New Brunswick’s (UNB) research department. OpenPhish is an automated web platform for identification and collection of phishing sites. URLs are submitted from various entities and networks across the globe for verification and OpenPhish use their in-built phishing detection scheme to validate the URL and extract various content from its body, like geolocations, brands, accounts, networks, etc. (openphish, 2019). The legitimate URLs were taken from UNB’s Canadian Institute of Technology’s research department, where they were exploring a lightweight approach to detect malicious URLs according to their attack vector and find if lexical analysis effective to identify these URLs. Benign URLs were taken from Alexa’s top websites by the Institute (Canadian Institute for Cybersecurity, 2019). The URL samples were taken at random to reduce any sort of inequity and decrease the outcome of unobserved factors.

### 4.1 Environment

Python, MS SQL server, and Android studio were used to carry out this research project. PyCharm IDE was used to write API scripts and create data visualization through python libraries (Lutz, 2013). An SQL server was preferred to save the API response over standard text file since it offers centralized data storage and recoverability. The prototype was tested on an Android Nexus emulator on API level 23 (Dawn Griffiths, 2017).

### 4.2 Evaluation criteria

To analyse and evaluate the performance of each API in terms of accuracy and latency, a confusion matrix is used for detailed understanding. Accuracy alone is not sufficient to justify the success or failure rate of an API if the dataset has an unequal set of observations (900:600). A confusion matrix is a table used to characterize the performance of the dataset (Saito, Rehmsmeier, 2015). The data is evaluated on a confusion matrix for the four outcomes of the binary classifier. The notations are detailed in Table 1.

**Table 1:** Confusion matrix

<i>N = Sample Size (1,500)</i>	<i>Predicted</i>		
		<i>Positive</i>	<i>Negative</i>
<b>Observed</b>	<i>Positive</i>	True Positive (TP)	False Negative (FN)
	<i>Negative</i>	False Positive (FP)	True Negative (TN)

To understand the evaluation criteria the following definitions are important:

- True Positive (TP): *The phishing URLs that are predicted correctly as phishing.*
- False Negative (FN): *The phishing URLs that are predicted incorrectly as legitimate.*
- True Negative (TN): *The legitimate URLs that are predicted correctly as legitimate.*
- False Positive (FP): *The legitimate URLs that are predicted incorrectly as phishing.*

In addition to plotting the confusion matrix, there are several parameters which are often calculated for a binary classifier for further analysis and evaluation (Saito, Rehmsmeier, 2015).

- Sensitivity (SN): Also known as Recall, is calculated as the number of correct positive predictions divided by the total number of positives. The best sensitivity is 1.0, whereas the worst is 0.0:

$$SN = TP / TP + FN$$

- Specificity (SP): It is calculated as the number of correct negative predictions divided by the total number of negatives. The best specificity is 1.0, whereas the worst is 0.0.

$$SP = TN / FN + TN$$

- Positive Predictive Value (PREC): It is calculated as the number of correct positives divided by the total number of positive predictions.

$$PREC = TP / TP + FP$$



- Negative Predictive Value (NPV): It is calculated as the number of correct negatives divided by the total number of negative predictions.

$$NPV = \frac{TN}{FN + TN}$$

- Accuracy: It is the overall probability of correctly classifying the URL (phishing or legitimate). Accuracy is calculated by:

$$\frac{TN + TP}{TP + TN + FP + FN}$$

- Error Rate or Misclassification (ERR): It is the overall probability of how often the dataset would result in an incorrect classification. It is calculated by the number of all incorrect predictions divided by the total number of records in the dataset.

$$ERR = \frac{FP + FN}{TP + TN + FP + FN}$$

Sensitivity is the probability of correctly predicting a phishing URL, when the URL is indeed a phishing. Specificity is the probability of correctly predicting a legitimate URL, when the URL is indeed a legitimate one. After plotting the confusion matrix, the above functions are derived and calculated from the data available from the matrix. Accuracy and Error rate are the most prevailing visceral measures evaluated from the confusion matrix.

## 5. General script pseudo-code for API request and response

The algorithm designed in this work is detailed below. The API keys/URL were defined in variables and then the dataset file was read in a loop. Each URL was sent to the API and response was collected and then saved in the database. This process was carried out for all the mentioned APIs.

```
Declare string API Key;
Declare string URL;
Open File (DataSet.txt);
While URL =0 to End of file do
    Response = Request.Check(URL);
    Display Response;
    Split Response into individual Key and Value;
    Establish Database connection;
    Insert into database (Individual response key, value objects);
End
Close Database connection;
Close File;
```

## 6. Experiment - analysis and results

### 6.1 Safe Browsing API

707 URLs were identified correctly from the total of 900 URLs. All 600 benign URLs were identified as legitimate. Google safe browsing provides 0 false positives, which means no legitimate URL is classified as a phishing one.

### 6.2 PhishTank API

PhishTank ensures that no legitimate URL is classified as malicious, but on the other hand, it falsely detects 457 malicious/phishing URLs as legitimate. Out of the 900 phishing URLs, 433 were identified correctly which is less than 50 percent of the total share. There were 457 False Negatives (more than 50 percent). The API had 100 percent success rate in identifying the true negatives and had zero false positives.

### 6.3 UrlScan.io API

675 URLs were identified correctly from the total of 900 URLs. It was also able to identify all the benign 599 URLs listed as legitimate, except 1, which was marked under False Positive. This URL was compared with the other 4 APIs and all marked it correctly as legitimate apart from UrlScan.io. Response time was the major issue endured in the experiment since the API being an automated process, which scans the entire website, collects its DOM objects and HTML content, the scanning process takes approximately 10 to 15 seconds.

### 6.4 CyberFish API

659 URLs were identified correctly from the total of 900 URLs. It was also able to identify all the benign 600 URLs listed as legitimate. Despite having computer vision and artificial intelligence integrated while verifying the URL, the response time of this API was much quicker than UrlScan.io and had a better accuracy than PhishTank.

### 6.5 Virus Total API

VirusTotal gave a 100 percent true positive rate since there was more than one engine to detect the phishing URL as a valid phishing link. Some antivirus engines can be aggressive in scanning and may falsely mark a URL as false positive for phishing. With an accuracy close to 100 percent, VirusTotal gave the best detection result amongst all the APIs. As VirusTotal interacts with 70 different anti-virus engines the response time taken to validate one request is between 10-15 seconds.

Figure 3 charts the API accuracy and response time results. Based on these two metrics, a comparative analysis can be carried out to identify which API would be the most suitable for mitigating smishing attacks based on the various possible end users' requirements.

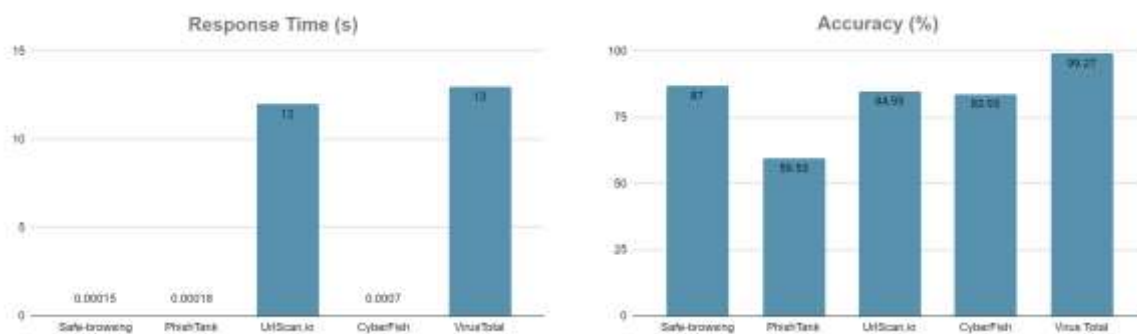


Figure 3: API response time and accuracy results

### 6.6 Response time

For an organization or individual whose major priority is response time while implementing any of the discussed APIs, the best option in this scenario is the Safe-Browsing API. With the response time of 0.15 milliseconds and accuracy over 87%, this API has the best accuracy and response time ratio. An alternative to this is CyberFish API which has a slightly higher response time at 0.70 milliseconds and an accuracy of 83.9%. A major drawback of CyberFish API is the number of credits available for each user. The free trial pack offers approximately 1000 credits, which equates to 1000 requests per account. If the limit is exceeded, then there is no way to verify the URLs unless the free trial pack is upgraded. In comparison, a client can make 10,000 requests per 24-hour period on one single API key for the Safe-Browsing API.

### 6.7 Accuracy

Accuracy can be an important metric for an organization while selecting an API. Even though the Safe-Browsing API offers 87% accuracy, it is still not the best possible fit available, exclusively in terms of accuracy. If the business case is more inclined towards detecting true positives, whilst having a high tolerance for latency, VirusTotal is the best possible fit available for integration. As discussed earlier, VirusTotal integrates 70 different anti-virus engines in its scanning process, the accuracy rate is approximately 99.27%, which is the highest amongst the five evaluated. A slight downside of this heavy integration is observed in the false positive rate. 11 URLs were marked as false positives due to the aggressive nature of the antivirus engines.

### 6.8 Filtering methodology

Organizations can prioritize the use of API based on the filtering methodology. Safe-Browsing maintains a repository of malicious URLs, but the repository is made with the combination of machine learning and URL extraction algorithms. CyberFish encompasses computer vision and artificial intelligence, whilst UrlScan.io works like an automated process and integrates external services for its results. PhishTank is completely different from the lot as it does not implement any working algorithm but is completely driven by the community and results

are declared depending on the vote share. Finally, VirusTotal incorporates different anti-virus engines in its interface providing a different perspective for phishing detection.

## **7. Conclusion and future work**

The objective of this research was to detect smishing attacks by integrating phishing detection APIs in the developed prototype and to evaluate five common and freely available APIs for performance and accuracy. Our results show that depending on different circumstances and business cases, different APIs can be more effective, answering the research questions posed at the beginning. From this research, it can be concluded that APIs are a good resource to be used in an application to detect smishing attacks and are helpful in making the system more secure and robust. The questions below which this research sought to answer are mostly in terms of how APIs can be used to detect phishing attacks and their integration in the prototype.

- 1. *Can a phishing detection APIs be used to detect smishing?* Our research shows that APIs can be used to detect smishing attacks and can be integrated in the prototype, however, not all APIs have the same accuracy or latency.
- 2. *Which phishing detection APIs are the most accurate and suitable for this prototype?* We observed that depending on business cases, developers can use different APIs to suit their requirements. For time sensitive applications, Safe-Browsing would provide the best solution, however, for security sensitive applications, VirusTotal would be a better option.
- 3. *Can API be integrated in the background process/broadcast receiver?* We successfully integrated the APIs in the broadcast receiver interface, which runs constantly in the background. An Alert message is displayed if the API detects a true positive.

The smishing detection prototype was built on the android operating system, however for universal use it can be expanded to iOS. To do this the prototype can be modelled in an ionic framework through which the application can be made compatible on more than one operating system. E.g., Android and iOS. Ionic is a new and upcoming open-source cross-platform framework used to build hybrid applications (Chaudhary, 2018). A limitation of our work includes the firewall configuration of some environments. Some premises may configure firewalls in a manner to block the API calls. In such a scenario, the mobile device is at risk for a smishing attack. Other limitations of the research as a whole were the number of freely available APIs. Paid APIs (commercial, enterprise) were not considered as part of the research.

Lastly, in order to scale the prototype as a full-fledged product, certain aspects need to be considered. To ensure integrity of the user phones, any application downloaded to the device must ask permissions for interacting with other applications. In addition to this, for the application to be used by a wider demographic, the application must be easy to use and compatible across various operating systems.

## **References**

- Chaudhary, P (2018). IONIC FRAMEWORK. International Research Journal of Engineering and Technology Canadian Institute for Cybersecurity. *url-2016.html*. Available at <https://www.unb.ca/cic/datasets/url-2016.html>. (2019, 06 01).
- Cyberfish. Available at <https://cyberfish.io/> (2019, 08 10)
- Developer.android.com. *broadcasts*. Available at developer.android.com: <https://developer.android.com/guide/components/broadcasts> (2019, 08 08).
- Developers.google.com. *Safe-Browsing*. Available at developers.google.com: <https://developers.google.com/safebrowsing/v4/> (2019, 08 05)
- Dhamija, R., Tygar, J.D. and Hearst, M. (2006). Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*
- Dawn Griffiths, D. G., (2017). *Head First Android Development: A Brain-Friendly Guide*.
- Hossain, S., Tulin, K. and Victor, C. (2015). Journal of Information Security. *Mobile Phishing Attacks and Mitigation Techniques*
- Jagatic, T.N., Johnson, N.A., Jakobsson, M. and Menczer, F. (2007). Social phishing. *Communications of the ACM*
- Longfei Wu, Xiaojiang Du, Jie Wu, (2014). *MobiFish: A lightweight anti-phishing scheme for mobile phones*.
- Lutz, M., (2013). *Learning Python: Powerful Object-Oriented Programming*.
- Ofoeda, J., Boateng, R. and Effah, J. (2019). Application programming interface (API) research: A review of the past to inform the future. *International Journal of Enterprise Information Systems*
- openphish. Available at <https://openphish.com/>(2019, 06 01).
- Pieterse, H., Olivier, M. and Heerden, R.V. (2019). Evaluation of smartphone data using a reference architecture. *International Journal of Electronic Security and Digital Forensics*.

**Pranav Phadke and Christina Thorpe**

- phishtank. *faq.php*. Available at phishtank.com: <https://www.phishtank.com/faq.php> (2019, 08 10).
- Shi, Y., Li, X., (2018). Computer Vision Imaging Based on Artificial Intelligence. *2018 International Conference on Virtual Reality and Intelligent Systems*.
- Saito T, Rehmsmeier M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets
- spamfighter. *News-6522-PhishTank-System--System-To-Catch-Phish.htm*. Available at <https://www.spamfighter.com/News-6522-PhishTank-System--SystemTo-Catch-Phish.htm> (2019, 08 10).
- Taylor, K. and Silver, L. (2019) 2019. *smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally* Available at <https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/>
- Urlscan.io. *about*. Available at <https://urlscan.io/about/> . (2019, 08 10)
- Virustotal.com. *reference*. Available at <https://developers.virustotal.com/reference>. (2019, 08 10).

# Evolving Satellite Control Challenges: The Arrival of Mega-Constellations and Potential Complications for Operational Cybersecurity

Carl Poole, Mark Reith and Robert Bettinger

Air Force Institute of Technology, Wright-Patterson AFB, USA

[Carl.Poole@afit.edu](mailto:Carl.Poole@afit.edu)

[Mark.Reith@afit.edu](mailto:Mark.Reith@afit.edu)

[Robert.Bettinger@afit.edu](mailto:Robert.Bettinger@afit.edu)

DOI: 10.34190/EWS.21.082

**Abstract:** The introduction of automated satellite control systems into a space mission environment historically dominated by human-in-the-loop operations will require the extraneous establishment of cybersecurity measures to ensure space system safety and security. With the addition and expansive growth of the “mega-constellation,” the old methods of satellite command and control are no longer cost effective. The proliferation of low Earth orbit (LEO) with thousands of satellites will require increasing levels of automation in order to handle internal operations, or the operations driven for the control of each constellation. The implementation of commercial off the shelf parts, coupled with on-board satellite computer systems that resemble the standard personal computer will allow for greater levels of automation and, therefore, fewer human interactions required to control newer satellites. On the ground segment side of satellite control, the influx of privately owned communication antennas for rent and a move to cloud-based operations or mission centers will present new requirements in cyber protection for both DoD and commercial satellite operations. This paper will highlight the changes these technical advancements will bring to the current satellite control architecture. It will also discuss likely ways that industry will evolve with the implementation of new requirements like the Cybersecurity Maturity Model Certification, finishing with a proposed change to how space and cyber space professionals can realign their interactions in order to address any emerging threats. It is no longer a matter of if automation will play a significant role in satellite operations, but how fast can the satellite operators adapt to the onset of control automation to promote cybersecurity in an increasingly competitive, contested, and congested space domain. One way for promoting such cybersecurity in space control is to introduce cybersecurity/monitoring training at all levels of satellite operations to align with the desire of creating a highly digitally-capable Space Force.

**Keywords:** satellite automation, mega-constellations, cybersecurity, ground station service, software defined equipment, future space operations

---

## 1. Introduction

The application of control automation has increased over the past two decades to the point that it has spread from industrial manufacturing and self-driving cars, to the home and household appliances. Control automation has also moved into the realm of satellite control operations. The focus in satellite control automation is driven on two fronts. First, the ability to incorporate cost effective, highly capable equipment in the spacecraft design allows for an increase in on-board controls processing. Second, the proliferation in orbital regimes – the focus for this paper will be on low Earth orbit (LEO) – is pushing complex tasks, such as satellite link scheduling and conjunction avoidance maneuvers, beyond the control of human operators. An additional operational distinction is made between satellite automation, or the self-contained process of conducting repetitive tasks, and satellite autonomy, which gives the satellite the ability to implement changes with limited to no human-in-the-loop (Hartley and Hughes, 1996). This distinction will add an additional level of complexity to the cybersecurity of satellite control.

Placing tasks previously controlled by humans under the control of a computer-executed algorithm may be the only viable way to manage the development of future mega-constellations and allow for realistic space traffic management (Butow et al., 2020). However, the prospect of improved space traffic safety and collision avoidance via control automation raises several concerns for consideration. While increasing the levels at which LEO constellations can interact and cooperate, the needed infrastructure and data exchange alterations will introduce new entry points on the cybersecurity front. The introduction of software-defined equipment, cloud-based mission control centers, and ground stations-as-a-service are prime examples. To address new potential cybersecurity issues in commercial space applications and automation at the lowest level, translating into quick response times, space and cybersecurity members will need increased interactive cooperation and mission understanding.

To help evaluate the interconnection between satellite control automation and cyber operations as it pertains to the Department of Defense (DoD), this paper will first provide a high-level view on how current operations are conducted and secured. Next, the future of mega-constellations and the requirement for control automation will be addressed. Additionally, an in-depth look into emerging satellite operations technologies and how they could affect the future of both U.S. Air Force (USAF) and U.S. Space Force (USSF) cybersecurity operations will be discussed. Finally, the paper will propose a cadre structure along with training changes that will address the importance of cooperation and synergistic operations between the space and cybersecurity career fields.

## **2. Current satellite control operations**

The control architecture for satellites has been developed and implemented virtually the same since its debut in the late 1950s. Starting with the launch of the first artificial satellites, each on-orbit system has mostly featured a unique design, function, and mode of operation. This uniqueness has led to self-contained and independent operating procedures controlled by the satellite owner. In the typical satellite control structure, information from the satellite – such as sensor or imaging data and state of health information – is downlinked during a scheduled time with the receiver. From the receiver, the information is processed and passed to the Satellite Operations Center. At the Satellite Operations Center, the data is reviewed for faults or required adjustments, and new instructions are planned. If a case arises where an orbital maneuver is required to correct for position or to change location (e.g., slewing, station-keeping, or collision avoidance with another object), one member of the operations team will script the commands for the prescribed maneuver. This script is then reviewed by several other team members before being passed to and processed by the human satellite operator. During the next scheduled uplink opportunity with the satellite, the commands will be sent from the Satellite Operations Center to the transceiver and then back up to the satellite for processing and command execution. This type of hands-on approach was needed due to constraints in the on-board systems, specifically limited computing power and proprietary operating structures. The emphasis on human control ostensibly meant reduced concerns for cybersecurity, as well as an increased sense of command situational awareness due to the human-use of protected ground communications systems and owner-controlled data links. Despite its benefits, this process can be very time consuming, with task scheduling becoming increasingly complex with the addition of new satellites to the satellite control architecture. Consequently, this human-in-the-loop satellite control architecture will be unable to effectively manage the size of mega-constellations of the near future.

## **3. Mega-Constellations will require automation**

The capability to develop constellations consisting of thousands of individual satellites controlled by one operator is no longer a wistful dream of science fiction. With the introduction of LEO constellations such as “Starlink” and “OneWeb,” the dream of mega-constellations has become a reality (McDowell, 2020). The proliferation of LEO with thousands of satellites will require increasing levels of automation in order to handle internal operations, or the operations driven for the control of each constellation, and to enable for future constellation growth in a given orbital altitude regime.

The creation of mega-constellations has come as a result of two factors. First, the shift in the commercial space industry to create standardized, rapid production, and high volume space-capable vehicles has caused both the size and cost of individual satellites to decrease drastically (Ben-Larbi et al., 2020). The ability to buy commercial-off-the-shelf parts, or COTS, instead of making proprietary hardware lowers the cost of research and development, thus accelerating system production. The second factor is a function of satellite size: as the satellites get smaller, then more satellites can fit inside the payload fairing of a single launch vehicle, which, in turn, drives down the cost per satellite to reach orbit. Overall, the costs of satellite design, production, and space launch are decreasing and allow for the proliferation of space to increase exponentially. Conversely, the costs associated with operations will unfortunately only increase if changes are not introduced to the current satellite control paradigm.

The evolution of satellite control from human-in-the-loop commands to automation will first require the mega-constellation, in concert with the ground communications networks, to de-conflict satellite pass times over system ground stations (Ben-Larbi et al., 2020; Bentley, 2017; Tominaga, Silva, and Ferreira, 2008). A pass time, by definition, is the time each satellite needs to downlink, or transmit, data to the ground antenna, and to uplink, or receive, commands from the Satellite Operations Center. Depending on the mission and amount of information transmitted, timing is critical. Also, access durations to each ground antenna are determined by orbital altitude: the lower the satellite altitude, the faster the satellite passes over a given point on the ground.

This planning will be increasingly important as the communication bandwidths become increasingly crowded due to more satellites flying within the ground receiver's view.

Since the start of the twenty-first century, an increase in CPU power has enabled the addition of programmable capabilities to on-board satellite sub-systems (Iovanov et al., 2003). A growing number of satellites are being equip with on-board systems that resemble a standard personal at home computer (Ben-Larbi et al., 2020, p. 12). This design architecture, in turn, increases reliability, and an on-board system can identify and correct for faults and adapt to changing parameters much faster than a human-in-the-loop system (Gilles, 2016). A human-in-the-loop system is comparatively slower due to data transmission and analysis delays, as well as the need for the human actions to be reviewed and verified prior to command uplink. One of the most common satellite control tasks is that of station-keeping, or maintaining a satellite in a predetermined orbital attitude and position. For mega-constellations, station-keeping could be accomplished with an autonomous attitude determination and control systems. With the increase in demand placed on the ground stations due to the vast number of contacts, the ability for each satellite to both determine and correct orbital attitude and position discrepancies will need to become autonomous (Thomassin, Ecochard, and Azema, 2017, pp. 2-7; Lee, Hwang, and Kim, 2008, pp. 2222-2225). However, shifting such attitude and orbit maintenance tasks away from the ground segment will require the introduction of a robust fault and error alert architecture to identify and notify the human satellite operators of any anomalous events. Ultimately, by raising more "house-keeping" commands into the purview of control automation will shift the satellite control work load from continuous hands-on, day-to-day human operations to an on-call human response control structure. The introduction of greater automation will also remove the likelihood of an incomplete command being sent by human satellite operators, or the need to check for unsafe commands prior to data uplink (Ben-Larbi et al., 2020).

#### **4. Satellite control evolution and why cybersecurity will be front-and-center**

While automation will play a large role in handling the satellite functions, the main changes for cybersecurity will come from the evolutionary shifts made in the ground control segments and associated security implementation requirements. In the 2020 USSF document "Space Capstone Publication: Spacepower Doctrine for Space Forces," the foundation for cyber security is defined in the Cyber Operations Spacepower discipline: "[The] knowledge to defend the global networks upon which military space power is vitally dependent; [the] ability to employ cybersecurity and cyber defense of critical space networks and systems; [and the] skill to employ future offensive capabilities" (Raymond, 2020, p. 52).

The future of security implementation is already being felt in the realm of manufacturing for DoD contracts. The recently introduced Cybersecurity Maturity Model Certification (CMMC) is intended to hold industry responsible for exercising sound cyber security practices, starting with the components and systems provided to the DoD by requiring it to use the published National Institute of Standards and Technology (NIST) rating system (Rosenberg, 2020). The CMMC is also rooted by the Federal Acquisition Regulation (FAR), Federal Information Processing Standards (FIPS), and general industry collaboration (CMMC-COE Administrator, 2020). The CMMC does have several caveats with one focusing on COTS systems (Rosenberg, 2020). This shift will ensure that the hardware and software being introduced for future satellite control needs will be primed for cyber defense. Another aspect that will play a role in the coming changes will focus around the protection of potential dual-use technologies. Butow (2020, p. 6) discusses that "entrepreneurs with innovative and potentially dual-use technologies must improve the protection of their intellectual property from unintended foreign assimilation, including protecting their networks from cyber exfiltration attempts, and avoiding exit strategies that transfer intellectual property to foreign control hostile to US interests." Some of these dual-use technologies can come in the form of software defined components that will allow for greater flexibilities in upgrading the on-orbit and ground control segments, especially in the area of communication systems (Manulis et al., 2020). Though software defined systems will add increased flexibility and allow for faster fixes if damaged (i.e., there is no need to replace expensive parts if the component can be simply reprogrammed), it will also introduce a new level of security requirements and response capabilities due to the inherent vulnerabilities in all software control systems.

Another area of evolving satellite control is related to the use of flexible ground control systems, more specifically the ground antennas used to send commands and receive data. Commercial entities such as Microsoft are introducing ground stations-as-a-service to increase capabilities and offset costs associated with satellite command and control (Hitchens, 2020). These systems will need to be very diverse in operational software and equipment to cover the wide range of satellite technologies currently used. Alternatively, future satellite

designs that intend to use this method of control can establish a form of technological standardization. In either case, commercializing this ground segment will be beneficial to handle the increased volume and bolster networked capabilities. Despite these benefits, network security for current satellites is based on such systems being owner controlled, system specific, and network isolated units. A new control structure is only half of the required change – the other half involves changing how and where some of the satellite control operations tasks are conducted.

This second change is coming in the form of cloud-based Satellite Operations Centers. As with the software defined component and commercialized ground stations, cloud-based control will allow for a more robust and flexible answer for growing constellations without the need to build costly new mission-specific “brick and mortar” centers (Ben-Larbi et al., 2020). This area already has several working examples, such as the “Major Tom” system, produced by Kubos (Kubos, 2020), which is implemented by the Planet company for use in its Dove constellation consisting of approximately 250 small satellites (Ben-Larbi et al., 2020). Cloud-based systems will have the added benefit of being accessible from any “secure” networked computer. In concert with the aforementioned commercial ground stations, mega-constellations could be truly operated from any location on the globe with a proper access point.

In the emerging satellite control dynamic, signals from a customer satellite would be sent to a configured service receiver. From there, it would be sent to the cloud-based Satellite Operations Center and accessed by satellite operations member from any networked system. Even with this control flexibility, the use of increasingly networked systems that are not owned directly by the operator can introduce new entry points and areas for cyber vulnerability (Scanlan et al., 2019). In order to ensure the cyber protection of all U.S. and Allied space-based assets, members directly in touch with these evolving systems will need to change just as drastically as the systems themselves (Department of Defense, 2020).

## **5. What a networked operations center could look like**

With the anticipated shifts in both the methods and infrastructure for space control operations, there will need to be an equal shift in the cadre structure and training for the operations teams. This will hold true for the DoD and commercial sectors alike. On the satellite operations floor, the phrase “a phone call is free” is often used when a member needs to reach out and address abnormal situations. However, this consultation only works well if the members on both ends of the call “talk the same language.” As the transition to increasingly networked centers interfacing with highly automated systems progresses, both the space operations professionals and the cybersecurity professionals will need to learn and understand more of the other members’ skill sets and technical terminology. Ideally, the formal training for satellite operations members will evolve to include space- and cyber-centric curriculum. This training could be in the form of introductory classes into cyber defense for the space professionals, and satellite mission design and communications for the cyber professionals. The USSF is in a crucial position to make this happen starting at the ground level, and the increased education will add to the understanding of “network dimension” (Raymond, 2020, p. 7). Optimally, this education would result in having embedded cyber operations members at key Satellite Operations Centers, in addition to having increased cybersecurity/monitoring training at all levels of satellite operations to align with the desire of creating a highly digitally-capable Space Force (Pope, 2020). Building a cyber-minded and space-proficient space control foundation will ensure that both the space and cyberspace cadre will have the tools needed to tackle any future growth in satellite capabilities and space mission execution. It will also empower members with the abilities and confidence to react rapidly and even preemptively to future threats.

## **6. Conclusion**

Satellite systems and controls architectures are in a rapid state of change. Satellite automation could significantly alter the current hands-on satellite operations mission to one of casual monitoring, with only sparse human-in-the-loop team present to react to and resolve issues that cannot be directly handed by the satellite itself. Additionally, the introduction of a more capable and increasingly flexible mission operations systems, one using emerging technologies such as cloud-based networks and services like that privately owned and networked ground stations, will make it possible for true 24/7 global access to satellite systems. In order to ensure the continued safety and security on-orbit satellite systems, both the defense and commercial space sectors must adapt to the rapidly changing digital landscape of future space operations. Such an adaptation is already demonstrated with the introduction of the CMMC initiation, along with the alignment of emerging USSF doctrine and strategy with cyber-mindedness. The final step will be to shape the future of the USSF and USAF space and



cyberspace cadre to be better prepared as a “digital” force synergistically working to remain at the forefront of protection in the increasingly competitive, contested, and congested domain of space.

*The views expressed are those of the author and do not reflect the official policy or position of the US Air Force, Department of Defense, or the US Government.*

## References

- Ben-Larbi, M.K., Pozo, K.F., Haylok, T., Choi, M., Grzesik, B., Haas, A., Krupke, D., Konstanski, H., Schaus, V., Fekete, S.P., Schurig, C., and Stoll, E. (2020) “Towards the Automated Operations of Large Distributed Satellite Systems. Part 1: Review and Paradigm Shifts,” *Advances in Space Research*, August, [online], <https://www.sciencedirect.com/science/article/pii/S0273117720305676>.
- Ben-Larbi, M.K., Pozo, K.F., Haylok, T., Choi, M., Grzesik, B., Haas, A., Krupke, D., Konstanski, H., Schaus, V., Fekete, S.P., Schurig, C., and Stoll, E. (2020) “Towards the Automated Operations of Large Distributed Satellite Systems. Part 2: Classifications and Tools,” *Advances in Space Research*, September, [online], <https://www.sciencedirect.com/science/article/pii/S0273117720305925>.
- Bentley, M.J., Lin, A.C., and Hodson, D.D. (2017) “Overcoming Challenges to Air Force Satellite Ground Control Automation,” 2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management, Savannah, GA, March, [online], <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7929585>.
- Butow, S.J., Cooley, T., Felt, E., and Mozer, J.B. (2020) “State of the Space Industrial Base 2020: A Time for Action to Sustain US Economic & Military Leadership in Space,” [online], <https://www.newspacenm.org/wp-content/uploads/2020/07/State-of-the-Space-Industrial-Base-2020-A-Time-for-Action-to-Sustain-US-Econ.-Mil-Leadership-in-Space.pdf>.
- CMMC-COE Administrator. (2020). “Understanding the CMMC Fundamentals,” Cmmc-Coe.Org, [online], <https://cmmc-coe.org/understanding-the-cmmc-fundamentals/> (Accessed: 18 January 2021).
- Department of Defense (2020) “2020 Defense Space Strategy Summary,” [online] [https://media.defense.gov/2020/Jun/17/2002317391/-1/-1/1/2020\\_DEFENSE\\_SPACE\\_STRATEGY\\_SUMMARY.PDF](https://media.defense.gov/2020/Jun/17/2002317391/-1/-1/1/2020_DEFENSE_SPACE_STRATEGY_SUMMARY.PDF).
- Gilles, K. (2016) “Flying Large Constellations Using Automation and Big Data,” SpaceOps 2016 Conference, Daejeon, South Korea, May, [online], <https://arc.aiaa.org/doi/10.2514/6.2016-2387>.
- Hartley, J.B. and Hughes, P.M. (1996) “Automation of Satellite Operations: Experiences and Future Directions at NASA GSFC,” Proceedings of the 1996 International Symposium on Space Mission Operations and Ground Data Systems, pp. 1262-1269.
- Hitchens, T. (2020) “Microsoft Boosts Space Services, Partnerships,” *Breaking Defense*, [online], <https://breakingdefense.com/2020/10/microsoft-boosts-space-services-partnerships/> (Accessed: 2 December 2020).
- Ivanov, M., Schulz, S., Dixon, G., Puderbaugh, A., and Shepperd, R. (2003) “Automation of Daily Tasks Necessary for the Management of a Large Satellite Constellation,” AIAA Space 2003 Conference & Exposition, Long Beach, CA, September, [online], <https://arc.aiaa.org/doi/abs/10.2514/6.2003-6209>.
- Kubos, “Major Tom,” [online], <https://www.kubos.com/majortom/> (Accessed: 30 November 2020).
- Lee, B.S., Hwang, Y., and Kim, H.Y. (2008) “Automation of the Flight Dynamics Operations for Low Earth Orbit Satellite Mission Control,” 2008 International Conference on Control, Automation and Systems, Seoul, South Korea, October, pp. 2222– 2225, doi: 10.1109/ICCAS.2008.4694468.
- Manulis, M., Bridges, C.P., Harrison, R., Sekar, V., and Davis, A. (2020) “Cyber Security in New Space,” *International Journal of Information Security*, May, [online], <https://link.springer.com/article/10.1007/s10207-020-00503-w>.
- McDowell, J.C. (2020) “The Low Earth Orbit Satellite Population and Impacts of the SpaceX Starlink Constellation,” *The Astrophysical Journal Letters*, Vol. 892, No. 2, pp. 1-18, doi: 10.3847/2041-8213/ab8016.
- Pope, C. (2020) “Driven by ‘a Tectonic Shift in Warfare’ Raymond Describes Space Force’s Achievements and Future,” SpaceForce.mil, [online], <https://www.spaceforce.mil/News/Article/2348423/driven-by-a-tectonic-shift-in-warfare-raymond-describes-space-forces-achievements/> (Accessed: 6 December 2020).
- Raymond, J. (2020) “Space Capstone Publication: Spacepower Doctrine for Space Forces,” [online], <https://www.spaceforce.mil/Portals/1/Space%20Capstone%20Publication%20Aug%202020.pdf>.
- Rosenberg, B. (2020) “‘Start Of A New Day’: DoD’s New Cybersecurity Regs Take Effect Today,” *Breaking Defense*, [online], <https://breakingdefense.com/2020/12/start-of-a-new-day-dods-new-cybersecurity-regs-take-effect-today/> (Accessed: 1 December 2020).
- Scanlan, J.D., Styles, J.M., Lyneham, D., and Lützhöft, M.H. (2019) “New Internet Satellite Constellations to Increase Cyber Risk in Ill-Prepared Industries,” 70th International Astronautical Congress, Washington, D.C., October, [online], <https://iafastro.directory/iac/paper/id/51953/summary/>.
- Thomassin, J., Ecochard, M., and Azema, G. (2017) “Predictive Autonomous Orbit Control Method for Low Earth Orbit Satellites,” *International Symposium on Space Flight Dynamics*, Matsuyama, Japan, June, pp. 2-7, [online], [https://issfd.org/ISSFD\\_2017/paper/ISTS-2017-d-086\\_ISSFD-2017-086.pdf](https://issfd.org/ISSFD_2017/paper/ISTS-2017-d-086_ISSFD-2017-086.pdf).
- Tominaga, J., da Silva, J.D.S., and Ferreira, M.G.V. (2008) “A Proposal for Implementing Automation in Satellite Control Planning,” SpaceOps 2008 Conference, Heidelberg, Germany, May, [online], <https://arc.aiaa.org/doi/10.2514/6.2008-3271>.



# **Work in Progress Papers**



# Inter-Process CFI for Peer/Reciprocal Monitoring in RISC-V-Based Binaries

Toyosi Oyinloye, Lee Speakman and Thaddeus Eze

University of Chester, UK

[t.oyinloye@chester.ac.uk](mailto:t.oyinloye@chester.ac.uk)

[l.speakman@chester.ac.uk](mailto:l.speakman@chester.ac.uk)

[t.eze@chester.ac.uk](mailto:t.eze@chester.ac.uk)

DOI: 10.34190/EWS.21.115

**Abstract:** Attacks stemming from software vulnerabilities that cause memory corruption often result in control flow hijacks and hold a place of notoriety in software exploitation. Attackers take advantage of vulnerabilities due to programming flaws to execute malicious code for redirecting the intended execution flow of applications. Existing defences offer limited protection due to their specificity to system architecture, operating systems or hardware requirements and are often circumvented by increasingly sophisticated attack techniques. This paper focuses on securing applications that are built on and run on the Reduced Instruction Set Computer Five (RISC-V *pronounced risk-five*) architecture, which is fast becoming popular on embedded devices such as smartphones, tablets, or other Internet of Things. Studies have revealed different threats that could emerge in an environment that is based on RISC-V architecture, drawing attention to growing demands for more resilient protections for RISC-V binaries. A concept based on Control Flow Integrity (CFI) appears to give promising solutions to control flow hijacks via various forms of implementation. The innovation in this research proposes an implementation of CFI with scrambled labels and logging of rogue attempts on vulnerable RISC-V-based applications. This would subsequently be extended for peer/reciprocal monitoring between similar binaries on RISC-V platforms.

**Keywords:** control flow integrity, RISC-V, buffer overflow, memory corruption, cybersecurity

---

## 1. Introduction

Memory corruption and Control Flow Hijacking (CFH) are notorious factors in software exploitation, as threats to the security of software has evolved from intrusions that are gained physically into attacks deployed via remote access. Memory bugs underlie vulnerabilities that are prevalent due to complexities in applications and operations that they are built to facilitate. Apart from memory corruption, insecurity in applications could stem from weaknesses due to underlying architectural structure of the system on which the binary is compiled and run. Existing protections built around X86 and ARM architectures are not adaptable to all platforms. Particularly, the RISC-V architecture, which is gaining popularity with producers of embedded devices, automotive systems, artificial intelligence, etc., are not adequately protected from CFH – redirection of the execution flow of a program. Furthermore, basic protections that aim to protect the memory have been circumvented through sophisticated attack techniques but protection that monitors the flow of execution via the CFI concept might give long lasting solutions. It has been implemented in various forms using hardware or software components to perform integrity checks on functions that are called by process prior to execution of critical instructions. Software-based CFI by Abadi et al., (2005) relies on Control Flow Graphs (CFG) that are obtained by conducting static analysis on binaries. CFGs are used for generating possible execution paths that is used to guide the monitoring process. Monitoring labels are inserted at critical function edges to enforce intended flow of processes at runtime. According to Payer (2017), CFI appears promising for mitigating control hijack tactics offering more precision and reliability. This study was inspired by the need for reliable protections for RISC-V based binaries. RISC-V is similar to the MIPS (Microprocessor without Interlocked Pipelined Stages) architecture and has small, highly optimised set of instructions that distinguishes it well from other more specialised sets that exist in other architectures. Particularly, adopting a load and store architecture with many registers that enable it to achieve an optimised instruction set, and a highly regular instruction pipeline resulting in low number of clock cycles per instruction (Chen, Novick and Shimano, 2000). In contribution to software protection measures, this study presents additional techniques based on CFI concepts to enhance protection for RISC-V-based binaries. Section 2 gives a literature review. Possible CFI solution from this Work-in-Progress is discussed and implemented in Section 3. Plans towards future works are discussed in Section 4 and Section 5 gives a conclusion to this paper.

## 2. Literature review

Valuable input has been made by researchers and software vendors towards software security, but attackers continue to advance ideas for circumventing former strong defences. Basic protections offer considerable

resistance to control flow hijacks, but over the years, exploits reveal lapses that are lurking within system facilities. Data Execution Prevention (DEP) eliminates code injections (Payer, 2020) but according to Göktas et al., (2014) sophisticated attack techniques like Return Oriented Programming (ROP) and return-to-libc (Buchanan et al., 2008) would bypass DEP.

Attackers usually begin their exploits gathering information and then proceed to inject payloads into memory spaces that are vulnerable. Address Space Layout randomisation (ASLR) (Pax Team, 2003) is effective for securing memory information but according to Shacham et al. (2004), it is vulnerable to information leakage. It is noteworthy to mention that these traditional protections are mostly useful in computers and servers but are limited in implementation on embedded devices that often run on the likes of RISC-V platforms as they do not use sophisticated memory management hardware, which are essential for implementing DEP and ASLR. In addition to this, RISC-V applications exist in highly sensitive eco systems as they are commonly used and constantly running. CFI is widely studied (Abadi et al., 2005; Payer, 2000; Neugschwandtner et al., 2016) and appears to offer promising solutions to CFH by mitigating further intrusion, even when attackers have gained full control of the application.

### **3. CFI solution**

An implementation of CFI on a simple vulnerable program is demonstrated here to show another possible way of implementing CFI for reliable protections, especially for RISC-V-based binaries.

#### **3.1 Methodology**

Initial in-depth reviews of existing works have informed the methodology and approach applied in this study which involves qualitative and quantitative research. A description of outcomes observed after running vulnerable programs with the new CFI technique would show how vulnerable programs are protected, while security and performance evaluation with the use of standard metrics would be applied to evaluate the new technique. The CFI concept is being evaluated with the aim of improving security of RISC-V binaries through its implementation. Experiments are conducted by implementing new CFI techniques and deliberately manipulating dependencies for running vulnerable programs in RISC-V environments.

##### *3.1.1 Experiment scenario*

The threat model in this study is an injection of malicious codes to a RISC-V-executable binary that accepts input from users and has a buffer overflow vulnerability. The simulation comprises a Linux Fedora host system where the command line is used for running a QEMU emulator on which a RISC-V Fedora Emulated Machine (EM) is booted, to write, compile and run the vulnerable binary in Figure 1 as depicted in Figure 2.

##### *3.1.2 Limitation*

All experiments are done using an emulator. Actual environments of embedded devices, automotive, artificial intelligence, etc would differ physically. Basic protections -ASLR and canaries were deliberately disabled to create desired threat model for this study and also to show how CFI could mitigate attacks that have circumvented these protections.

### **3.2 Possible solution**

A solution being proffered here adopts CFI by inserting checks and enforcements at critical edges within functions that exist in vulnerable binaries and then taking a log of every redirection attempt that was subverted. The log is valuable for tracking rogue functions specifically to identify whether it was a reused code or a code injection.

##### *3.2.1 Implementation*

Static and dynamic analyses were done on a vulnerable binary to map out its CFG. The source code was then recompiled into assembly code where additional lines of instructions were inserted to perform desired checks and enforcement. The checks were implemented as shown in Figure 3 by applying label value that corresponds at all edges for all functions that were intended to be executed within the program. An additional unintended function *func2* was included in the vulnerable program and assigned a different label values to demonstrate redirection of execution flow by using a buffer overflow to overwrite the return address in *func1* with the address

of *func2*. Labels were checked by comparing them with a monitor that resides in the main function (caller) just before the call instruction to *func1* (callee).

```
#include <stdin.h>
#include <string.h>
#include <stdlib.h>
#include <sys/time.h>

#define MAX_NAME_SIZE 256

void func1(void)
{
    int number = 0; // over-flow target
    char buffer[16]; // to hold the player's name

    printf("Please enter your name: ");
    fgets(buffer, MAX_NAME_SIZE, stdin); // input from "Standard In"

    printf("buffer address: %x\n", (unsigned int)buffer);
    printf("number address: %x\n", (unsigned int)&number);

    printf("Welcome ");
    printf(buffer);
    printf("The number is %d\n", number);

    return;
}

void func2(void)
{
    printf("This function never gets called!\n");
    exit(-1);
}

int main(void)
{
    // Program starts here.
    char some_space[80];
    printf("buffer address: %x\n", (unsigned int)some_space);
    func1();

    return 0;
}
```

Figure 1: Source code of vulnerable program



Figure 2: Shows how systems are inter-related in the experiment scenario

In this example, the checks were inserted at two locations within each function that could be called in the program. First, at the function epilogue and then at the function prologue. The second insertion is useful for tracking down code executions due to jumps that could be made directly to instructions that reside after the start of the function, thereby bypassing the first check in the unintended function. A function *funclogger* is also inserted to take a log and direct flow to another function *Wrongmove* which triggers a system call to halt the program.

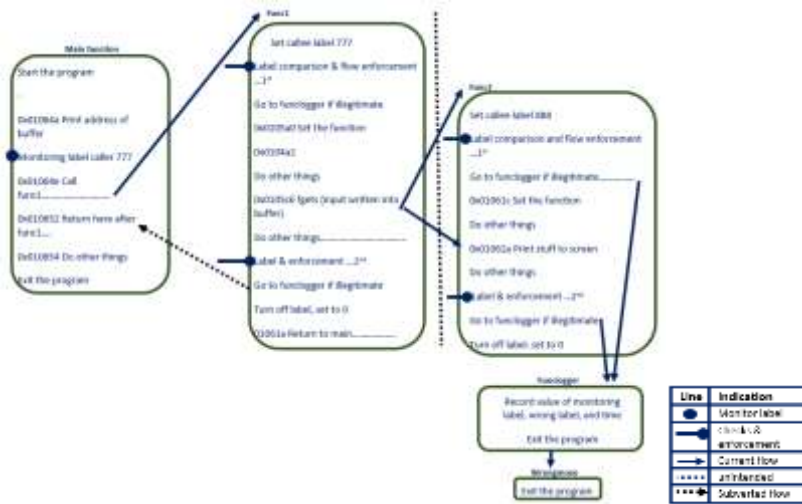


Figure 3: Showing CFI checks and enforcements

The instruction for assigning a monitoring value 777 which corresponds to that of intended *func1* is highlighted in the assembly code of the caller shown in Figure 4.

```
main:
    addi    sp, sp, -96
    sd     ra, 88(sp)
    sd     s0, 80(sp)
    addi    s0, sp, 96
    addi    a5, s0, -96
    sext.w  a5, a5
    mv     a1, a5
    lui    a5, %hi(.LC1)
    addi    a0, a5, %lo(.LC1)
    call   printf
    li     s3, 777
    call   func1
    li     a5, 0
    mv     a0, a5
    ld     ra, 88(sp)
    ld     s0, 80(sp)
    addi    sp, sp, 96
    jr     ra
```

Figure 4: Main function with instruction to assign monitoring value highlighted

Assembly code for unintended *func2* is shown in Figure 5 highlighting where a different label 888 is assigned to it followed by checks and enforcements.

```
func2:
    li     s4, 888
    bne   s4, s3, funclogger
    addi    sp, sp, -16
    sd     ra, 8(sp)
    sd     s0, 0(sp)
    addi    s0, sp, 16
    lui    a5, %hi(.LC5)
    addi    a0, a5, %lo(.LC5)
    call   puts
    li     a0, -1
    bne   s4, s3, funclogger
    li     s4, 0
    call   exit
```

Figure 5: Func2 including integrity checks and enforcement

With the CFI in place, the assembly code was then compiled into an executable binary which was run in gdb. A disassembly was done to retrieve the address of *func2*. Figure 6 shows where the vulnerability lies within *func1*.

```
(gdb) p &buffer
$2 = (char (*)[16]) 0x3fffffff148
(gdb) x /20xw $sp
0x3fffffff140: 0xffffffff358 0x00000003f 0x000000000 0x000000000
0x3fffffff150: 0xffffffff1f8 0x00000003f 0x000100000 0x000000000
0x3fffffff160: 0xffffffff1d0 0x00000003f 0x00010652 0x000000000
0x3fffffff170: 0xffffffff298 0x00000003f 0xf7fec18e 0x00000003f
0x3fffffff180: 0xf7fc5ce0 0x00000003f 0xf7ed6d98 0x00000003f
```

Beginning of the 16-byte buffer. This buffer could be overrun with some input to overwrite content of the cells that occur immediately after it.

Return address can also be overwritten.

Figure 6: Regions of memory that could be overrun

A carefully crafted input was supplied and gets written to the buffer as shown in Figure 7, to overwrite the intended return address and redirect the flow to *func2*.

```
(gdb) x /20xw $sp
0x3fffffff140: 0xffffffff358 0x00000003f 0x41414141 0x41414141
0x3fffffff150: 0x41414141 0x41414141 0x42424242 0x42424242
0x3fffffff160: 0x42424242 0x42424242 0x0001061c 0x000000000
0x3fffffff170: 0xffff000a 0x00000003f 0xf7fec18e 0x00000003f
0x3fffffff180: 0xf7fc5ce0 0x00000003f 0xf7ed6d98 0x00000003f
```

Buffer is filled with some junk followed by the redirection address

Figure 7: Input written to buffer overwrites return address



Once a call hits the epilogue of *func2*, CFI tracks down the redirection and triggers *funclogger* to take a log of the event and then transfer flow to function *Wrongmove* to halt the process.

### 3.3 Performance evaluation

Performance overhead would depend on the % per control flow checks in each program. Limitations observed in this implementation exist in form of possible information leakage of the monitoring value and increased execution time with additional checks insertion into the vulnerable program. A possible solution to leakage of monitoring values is to scramble the values and load them to appropriate registers at runtime.

### 4. Future works

Further works on this research will be creating code for auto-generating and scrambling of label values at each runtime. This can be achieved by using system clock values to form an integer at each program runtime. Once this is achieved, the new CFI enforcement will be extended to similar programs and applied for peer/reciprocal monitoring between processes. Peer monitoring would involve an independent, isolated master process that serves as monitor for other related processes as shown in Figure 8, while reciprocal monitoring would involve two independent and isolated processes reciprocating monitoring between each other while the master monitors other vulnerable processes as shown in Figure 9.

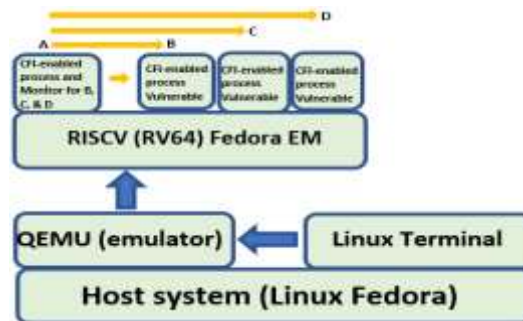


Figure 8: Peer monitoring for CFI-enabled processes

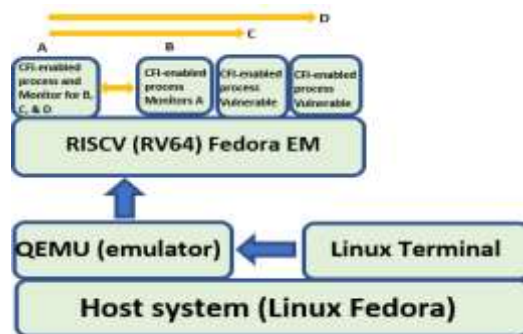


Figure 9: Reciprocal monitoring for CFI-enabled processes

### 5. Conclusion

This Work-in-Progress has explored insecurities in software demonstrating a means of mitigating CFH by implementing CFI for RISC-V based applications.

### References

Abadi, M., Budiu, M., Erlingsson, U., and Ligatti, J., 2005. *Control Flow Integrity*. NY, USA, Association for Computing Machinery, pp. 340-353.

Buchanan, E., Roemer, R., Shacham, H. & Savage, S., 2008. When Good Instructions Go Bad: Generalizing Return-Oriented Programming to RISC. *Proceedings of the 15th ACM Conference on Computer and Communications Security*, pp. 27-38.

Chen, C., Novick, G., and Shimano, K., 2000. *csstanford.edu*. [Online] Available at: <https://cs.stanford.edu/people/eroberts/courses/soco/projects/risc/riscisc/> [Accessed 21/05/21].

Göktaş, E., Athanasopoulos, E., Bos, H., and Portokalidis, G., 2014. *Out Of Control: Overcoming Control-Flow Integrity*. San Jose, CA, USA, IEEE, pp. 575 - 589.

**Toyosi Oyinloye, Lee Speakman and Thaddeus Eze**

- Neugschwandtner, M., Mulliner, C., Robertson, W. & Kirda, E., 2016. Runtime Integrity Checking for Exploit Mitigation on Lightweight Embedded Devices. *International Conference on Trust and Trustworthy Computing*, August, Volume 9824, pp. 60-81.
- Pax Team, 2003. *ASLR documentation*. [Online] Available at: <https://pax.grsecurity.net/docs/aslr.txt> [Accessed 22/04/20].
- Payer, M., 2017. Control-Flow Hijacking: Are We Making Progress?. *ASIA CCS '17*, 02-06 04.
- Shacham, H., Page, M., Pfaff, B., Goh, E-J., Modadugu, N., and Boneh, D., 2004. On the Effectiveness of Address-Space Randomization. *Proceedings of the 11th ACM Conference on Computer and Communications Security*, p. 298–307.

# Use of Blockchain Technologies Within the Creative Industry to Combat Fraud in the Production and (Re)Sale of Collectibles

Alexander Pfeiffer<sup>1, 2, 3</sup>, Stephen Bezzina<sup>2, 3</sup> and Thomas Wernbacher<sup>1</sup>

<sup>1</sup>Center for Applied Game Studies, Donau-Universität Krems (DUK), Austria

<sup>2</sup>Department of Artificial Intelligence, University of Malta (UoM), Msida, Malta

<sup>3</sup>B&P Emerging Technologies Consultancy Lab Ltd., St. Julian's, Malta

[Alexander.pfeiffer@donau-uni.ac.at](mailto:Alexander.pfeiffer@donau-uni.ac.at)

[mail@stephenbezzina.com](mailto:mail@stephenbezzina.com)

[Thomas.wernbacher@donau-uni.ac.at](mailto:Thomas.wernbacher@donau-uni.ac.at)

DOI: 10.34190/EWS.21.055

**Abstract:** The music industry has evolved significantly over the last few decades, from cassette to compact disk to MP3 and now to subscription-based streaming. Simultaneously, there has been a return to analogue, especially to vinyl records. In 2021, a major record label will introduce a new kind of vinyl. From the original master tapes, one-of-a-kind copies will be made. These will be manufactured in very limited quantities and sold exclusively as collectors' items. In a world where purchasing these collectibles is as simple as tapping the screen and where there are also numerous trading markets between private individuals, new creative ways to protect consumers and digitally protected analogue collectibles must be found. This relates to both the product's authenticity and the legitimate possession of the valuable vinyl. This work in progress paper aims to determine whether digital identities of suppliers, distributors, and consumers on the one hand, and decentralized encrypted data storage on the other, can be potentially the future technology to safeguard collectibles that the creative industry should be more than just looking at.

**Keywords:** collectibles, blockchain, digital ID, vinyl

---

## 1. Introduction to the topic

Collecting not only for survival but for cultural or educational reasons has accompanied mankind for thousands of years (Wilde, 2015). The very first collections of cultural goods as burial artefacts go back to 3000 BC in ancient Egypt. Early collections of books already existed in the ancient world by Polycrates of Samos and Euripides. Since the 6th century AD, collecting has been the privilege of the European ruling houses and the churches in their treasuries. This demonstrated power, wealth and influence (see *ibid*). For the first time, in the Renaissance, tendencies towards individualization become noticeable, and with it begins the golden age of collecting luxury goods that serve the "knowledge of the world". The first private collections emerged in the 14th century, and from 1450 onward, specific rarities and precious objects were collected, such as antique manuscripts, coins, statues, but also plants and the first cabinets of curiosities were created. In the following period of the Enlightenment, the society becomes modernized, and consequently collecting is socialized and democratized and enters the classical bourgeoisie. Two types of collection, the public and the privately owned, were established side by side (Werner, 2018). In the private sector, a new type of collection emerged in the 19th century - the small collection. Its purpose is to give cultural and intellectual pleasure to the collectors and add value to social life. The private collection is often an expression of the personality and identity of the owners (Wilde, 2015).

In the 21st century, the trend towards private collections has manifested itself. Research shows that every third German has a collecting impulse. For Austria, with its population of just under 10 million, a country that is culturally close to Germany, it is assumed that three million people collect items privately (see Schindelbeck, 1997, Sommer 2011, Jolmes, 2014). The most diverse things are collected; from curiosities, antiques, collectible stickers to music. Also, in the 21st century, collectables increasingly include digital items, such as computer games that were installed without data carriers, items as part of computer games or music songs and albums in their pure digital form or access to a streaming service (Werner, 2018, Pfeiffer, 2018).

In regard to Blockchain technologies Serada et. al. (2020) analyzes specific characteristics of value created through digital scarcity and Blockchain-proven ownership in cryptogames. Although their research relates to a digital collection game, conclusions can be well drawn for our project. Pfeiffer et. al. (2020) have also examined Blockchain technologies and found that it is precisely the safeguarding of collective objects, in other words digital items of real-world value that can be optimally secured by Blockchain. The connection between real-world

objects and Blockchain technologies for marking ownership and ensuring that it is an original is very well illustrated in the developer docs of the company Riktig. These provide an important resource for our work.<sup>1</sup>

## **2. Planned research aims and methods used**

This work in progress / case study paper deals in particular with the current trend of collecting vinyls<sup>2</sup>, specifically within the conceptualization of a major record label to produce strictly limited special editions directly from the original master tape. Besides the new complex manufacturing process and the strict limited availability, the aspect of counterfeit protection on the one hand and the proof of ownership on the other hand are of utmost importance. In addition, new business models are to be considered. For this purpose, the use of Blockchain technologies, specifically a demonstrator on the Blockchain Ardor/Ignis<sup>3</sup> is being developed. The first iteration of this demonstrator is described below. This served as the basis for the expert opinions in the frame of this study.

The research questions for the final paper are divided in two parts. In the first quantitative study, collectors of vinyls will be surveyed regarding:

- their overall motives;
- why this medium is so attractive to them as a collectible;
- how they see the future of the collecting of vinyls;
- whether the protection and certification of an authentic item plays a role for them;
- and finally, if the link to a digital identity and the connection to a customer to customer (C2C) sales platform handled via smart contracts could offer advantages.

The second part will be conducted in the form of expert interviews. In this work in progress, the key statements of the first interviews are presented, which have served to find the questions for the upcoming planned longer series of expert interviews. The focus here is on the issue of counterfeit protection, allocation to identities and the establishment of a C2(B)C marketplace. The latter means that customer sells to customer, but companies are integrated as part of the automated smart contracts via Blockchain. This present two key advantages, mainly for identity verification of the traded object and its owner on the one hand and on the other hand in the form of receiving a brokerage commission as part of a new business model.

## **3. Presentation of the developed demonstrator**

As part of the demonstrator, restrictive utility tokens are created on Ignis, the Blockchain Ardor's childchain, which will act as the digital counterpart of the respective limited edition vinyl. Restrictive means that the conditions under which the utility token can be sent from one wallet to another are defined in the form of approval models. For example, wallet addresses can be whitelisted, multi-account signature processes can be defined, or it can be determined that the possession of an authorization token is necessary to initiate certain actions.

Two different types of tokens are generated. The first type of tokens are singleton tokens (NFTs). These are unique tokens of which exactly 1 piece exists and which have their own transaction ID. This would be ideal for particularly valuable unique pieces. Here, the token description can be used for a general publicly visible description of the token, for example, which real good is digitally represented including, for instance the serial number and other factors which describe the product in general and show its unique characteristics. For each transaction, encrypted or unencrypted messages can be used to store meta-data.

This possibility is essential for our demonstrator, because the digital signature that identifies the owner is written into this as attached message, stored on Blockchain, accessible in form of the unique transaction ID. The second token corresponds to the number of pieces of a planned edition. Here the token description is the same for all tokens. If for example, a limited edition of 2000 pieces is produced, the same amount of tokens which share the same token ID and a common public description of the token properties are generated. The individualization,

---

<sup>1</sup> See <https://riktig.io/docs/developer>

<sup>2</sup> See <https://www.catawiki.de/stories/735-warum-man-jetzt-mit-dem-sammeln-von-vinyl-anfangen-sollte>

<sup>3</sup> See <https://www.jelurida.com>

such as the serial number, which vinyl it is, the issuer, the year of issue, the owner and other relevant meta data is therefore attached as an encrypted message to each transaction.

Using the marketplace of the Blockchain Ardor, the sales process from one account to another is simulated in the form of a role play. This involves the transfer of ownership rights, payment processing directly with cryptocurrencies or via digitally signed instructions to the house bank and the possibility of matching with databases of the manufacturer for additional verification of the vinyl, including update of the respective legal owner. We will also work out how to implement concepts where the system records and verifies private data (including ownership) but no unauthorized people can read this data. As such, future research should be directed towards what is commonly referred to as zero knowledge proofs, whereas one party can prove to another that they know a value  $x$ , without conveying any information apart from the fact that they know the value  $x$ .

#### **4. Preliminary expert statements**

In this Work in Progress / Case Study Paper, we would now like to sum up briefly the results from the first round of expert interviews. The statements refer on the one hand to the feedback on the first iteration of the demonstrator and on the other hand to essential points, which have to be considered for the completion of the full paper and the development of the questionnaire for the collectors. For this work in progress / Case Study we refer to statements from three selected experts

E1 is a tech-savvy musician who also manages his band. The amateur band usually plays in front of audiences of around 150 people and sell about 500 CDs per edition. Furthermore, they have a few thousand counted streams of their songs per year. The creation of a small edition vinyls for the closest fans has already been considered.

Even though it's probably not a big issue for his band yet, he finds the collector's edition aspect exciting. For his fans, this is also a bit of gambling on the band's future success. For him, it is important that he, as a small self-publisher, gets access to the use of the systems, via for example, a licensing system of external providers (such as the vinyl or CD pressing plant be). Also, he will get the possibility, after a review of certain factors, to register his own label with large major labels and sell the limited editions, as well as to use the new resources, such as the online store, for collecting such special editions.

E2 is not from the music industry, but an expert in Blockchain based processes and smart contracts. He finds the demonstrator very exciting and the concept well developed. He suggests thinking about the possibility of multi-chain connections early on, so that different systems and Blockchains can communicate with each other. He also suggested a design where the private information is mapped separately from that of the collection piece, but nevertheless through approval models one part does not work without the other. In this way, current European data protection law can be better addressed, such that private information remains private. For him, it is important to look at the transaction costs and who will be responsible for them. An important issue is the area of digital identities, and here it is particularly important to think beyond national borders.

E3 is the executive director of the major label releasing the novel collector's editions. While the launch of this collection is not yet secured on Blockchain, there is interest in the technology and this independent accompanying research serves as an initial evaluation. While the rising consumption of streams has become the backbone of the recorded music economy, the vinyl is the only format that defies the decline of the physical business. The reason for that is the iconic history and emotional value of the vinyl disc and the fact that it is strictly analogue i.e. anti-digital. The question is, how a very digital and modern concept as Blockchain sits with the hardcore vinyl collectors and how the value of such an evidence of ownership can be communicated. Another requirement would be the easy usability for both the manufacturer of the limited editions, as well as the owners.

#### **5. Conclusion**

The initial expert interviews provided a sound basis for the development of the quantitative questionnaire. Further steps can also be taken for the second iteration of the demonstrator, which will then be discussed in the next expert round, where we plan to discuss the results of the online survey together with the second iteration of the demonstrator. With regard to the development of the demonstrator further possibilities offered by the Ignis Blockchain will be explored in further iterations of the demonstrator, such as the division into 2 token systems for the same collectible. One token will contain the private owner data (encrypted) and another token

the public data around the vinyl. Both tokens can only be sent together, which is guaranteed by the corresponding approval model. The work so far shows that the topic of combining Blockchain-based digital identities with tangible collector editions may be novel but very exciting for both the industry and the end user, namely the music fan and collector. Finally, such findings are potentially of interest for applications within other creative industries.

## References

- Ilgen, V., Schindelbeck, D. (1979) Die Jagd auf den Sarotti-Mohr, Fischer Taschenbuch Verlag, Frankfurt, 1997
- Kleine, J., Jolmes, M. (2014) Sammeln : Im Spannungsfeld zwischen Leidenschaft und Kapitalanlage, Steinbeis Research Center for Financial Services, München
- Pfeiffer, A. (2019). Doktorat Alexander Pfeiffer: Auf dem Weg zur ludischen Gesellschaft. 10.13140/RG.2.2.21808.30725.
- Pfeiffer, A., Kriglstein, S., Wernbacher, T. (2020) Blockchain Technologies and Games: A Proper Match? In International Conference on the Foundations of Digital Games (FDG '20). Association for Computing Machinery, New York, NY, USA, Article 71, 1–4. DOI: <https://doi.org/10.1145/3402942.3402996>
- Serada, A., Sihvonen, T., & Harviainen, J. T. (2020). CryptoKitties and the New Ludic Economy: How Blockchain Introduces Value, Ownership, and Scarcity in Digital Gaming. Games and Culture. <https://doi.org/10.1177/1555412019898305>
- Sommer, M. (2011) Eine Phänomenologie des Sammelns und des Sammlers, in: Kunstforum international, Band 211, Ruppichteroth
- Werner, S. (2018) Die ungebrochene Faszination von Sammelbildern in einer digitalen Welt, Donau-Universität, Krems
- Wilde, D. (2015) Dinge sammeln, Annäherungen an eine Kulturtechnik, transcript Verlag, Bielefeld

# Peer2Peer Communication via Testnet Systems of Blockchain Networks: A new Playground for Cyberterrorists?

Alexander Pfeiffer<sup>1, 2, 3</sup>, Thomas Wernbacher<sup>1</sup> and Stephen Bezzina<sup>2, 3</sup>

<sup>1</sup>Center for Applied Game Studies, Donau-Universität Krems (DUK), Austria

<sup>2</sup>Department of Artificial Intelligence, University of Malta (UoM), Msida, Malta

<sup>3</sup>B&P Emerging Technologies Consultancy Lab Ltd., St. Julian's, Malta

[Alexander.pfeiffer@donau-uni.ac.at](mailto:Alexander.pfeiffer@donau-uni.ac.at)

[Thomas.wernbacher@donau-uni.ac.at](mailto:Thomas.wernbacher@donau-uni.ac.at)

[mail@stephenbezzina.com](mailto:mail@stephenbezzina.com)

DOI: 10.34190/EWS.21.049

**Abstract:** Peer2Peer communication can take place in the traditional way via e-mail, forums or social media. One also finds dedicated apps for communication or organized in groups, such as WhatsApp, Telegram or Discord, the latter being particularly popular with digital gamers. Online games are another medium which can foster communication between people over a data connection, as direct messages can be sent through the provisions of the digital game worlds. Depending on the game provider and its headquarters, the terms and conditions differ in how the data is transmitted and processed. Access to private communications is important for governments and especially for the police work, for both to prevent and follow up on cybercrime and terrorist acts. On the other hand, the private and civil rights movements push for such interventions to occur only in the case of absolutely justified suspicion, with otherwise restricted access to transmitted conversations and data of private individuals and companies. Therefore, it is important that such access to messages is confirmed in advance by a law court. But even with approval, it is still difficult for the authorities to gain access from a technical perspective. While IP addresses and open communication can be intercepted quite easily, it is more difficult when secure messenger apps are used and only possible if there is direct access to the user's device or the app operator provides the authorities access via a master key. In digital games, access is even more complicated. In this work-in-progress paper the authors want to address a currently overlooked aspect of Peer2Peer communication; which is the provision of text messages via (testnet) blockchain systems, with special regard to the possibility of attaching encrypted messages to the transaction of blockchain tokens. It is to be noted that on the testnet versions of the blockchain systems no "KYC" takes place. While on the mainnet versions of the blockchain systems the purchase of tokens to send them later can only be done anonymously "over the counter", the testnet of most blockchain systems is completely free available. Everyone can create a blockchain Wallet, request testnet tokens and start sending encrypted messages anonymously. This work-in-progress paper aims to highlight and explain the authors' planned research in this field.

**Keywords:** blockchain, DLT, social media, utility tokens, cryptocurrencies, rewards

---

## 1. Introduction to the topic

The issue of communication between individuals, cells or gangs with criminal intentions using modern communication technologies has been part of the ongoing debate for a long time. Not only since 9/11 the issue of monitoring the communication between citizens is being discussed and investigated, but above all when is surveillance legitimate. Various preventive measures against crime have been taken, such as the self-registration obligation for pre-paid SIM cards, as is currently the case in the EU<sup>1</sup>. However, in reality, as the white paper "The Mandatory Registration of Prepaid SIM Card Users" from 2013 suggests:

*"An increasing number of governments have recently introduced mandatory registration of prepaid SIM card users, hoping that the policy would support law enforcement and counter-terrorism efforts. However, to date there is no evidence that mandatory registration leads to a reduction in crime."*

It should also be noted, though, that this publication was written by a group representing the interests of the mobile communications industry.

The measures taken by governments in the fight against terrorism and the capabilities to read and analyze conversations go one step beyond the mere registration of devices and sim cards. At the end of 2020, for example, a cabinet decision was taken in Germany to be able to read encrypted messages from WhatsApp or

---

<sup>1</sup> Cf. Registration Law in Austria, according to EU legislation: <https://www.bmlrt.gv.at/english/telecommunications-and-postal-services/telecommunications-/registration-of-mobile-phone-prepaid-card.html>, last Accessed 26.01.2021

Facebook Messenger<sup>2</sup>, although it was emphasized that this is allowed only in individual cases and after a court order. To access these messages, however, it is necessary either to install malware / Trojan horses on the user's devices, or to gain direct personal access to the devices for a short period of time. In this case, the browser services of the chat platforms are used by the police investigators, who connect the cell phone of the suspect to the browser service via a QR code and can thus read the messages<sup>3</sup>.

Another way to communicate in an encrypted modality is via email and OpenPGP | S/MIME. In 2018, however, the research group behind efail published how, under certain conditions, but mostly due to user errors in the use of the encryption service, access to the plaintext of messages can be obtained. Another potential area of non-supervised communication is video games. In 2011, Thomas Gabriel-Rüdiger and Cindy Krems, both cybercrime experts from Germany, gave a lecture at the Danube University Krems in which they showed how online role-playing games can be used to plan terrorist activities<sup>4</sup>. It is especially difficult to track down when the language code is based on in-game terms or otherwise. For example, the name of a boss monster is assigned to the terror target. So the cell pretends to be planning a raid. Coordinates from the game can also be used here, which can then be applied to maps in the real world. Such conversations are even very difficult to interpret for Artificial Intelligence algorithms. Especially when working with private chats, it is also impossible for a gamemaster to notice. And so, these conversations can only be intercepted if there is already a suspect of terrorism, the user name of the player is known, and there is a corresponding court order that allows the authorities to cooperate with the game producers to receive chat data in real time.

In 2015, a broader discussion on this topic began, especially through mass media. This has partially forced the manufacturers to adjust the Terms and Conditions, and in some games players now agree that chat messages are stored for a certain time. As an example of the media discussion, a debate on this topic can be found on CBSN<sup>5</sup>. A large audience was made aware of this possibility of communication via a fictional terror scenario. In the television series Jack Ryan (2018), a (not in real-life existing) computer game was shown via which terrorists communicate across countries. What is particularly exciting is that this example encourages us to think not only of large commercial games, but of independent games, i.e. low-budget games with encrypted voice or text chats; games that can be used just under the radar of the investigators. Another important aspect in the discussion of using games as methods of communication is the aspect of data ownership. The senders of the messages do not know, apart from the set of rules described in the terms and conditions, what actually happens with their messages. For example, how long they are stored, and which authorities get access rights at which point in time. An alternative would be dedicated computer games that are used as a cover, but once revealed, these would be a fish pond for the investigators and therefore it is not assumed that these approaches actually exist.

The authors will now take a look at a communication technology that still seems unnoticed by cyberterrorism academia, as well as by the authorities. Blockchain and especially testnet systems of blockchain networks as means of communication. By definition, a Blockchain is a continuously growing chain of blocks, each of which contains a cryptographic hash of the previous block, a time-stamp, and its conveyed data (Nofer et. al, 2017). Grech and Camilleri (2017) describe (positive) effects of Blockchain technologies, like self-sovereignty, trust, transparency, immutability, disintermediation and collaboration. The concept of Blockchain, as we know it today, derives from Satoshi Nakamoto's Whitepaper 'Bitcoin: A Peer-to-Peer Electronic Cash System', published in late 2008. Originally intended to create a non-manipulable account book to represent the possession of digital tokens, which in turn are traded for money on exchanges or over-the-counter (peer2peer), it is now about the technology behind it and what applications can possibly be developed using Blockchain technology to secure transactions. The idea of using the Bitcoin Blockchain for more than 'proof of payment transactions' arose from the fact that you can attach text messages to a transaction. To create an account book of any imaginable transaction, a fraction of Bitcoin (so-called Satoshis) was sent to an address and the text to be recorded was attached to it as a text message and thus stored forever on Blockchain. However, if such information is simply stored as a text message attached to the same kind of token, this strongly limits its possible applications. And since Bitcoin was not originally intended for other applications apart from payment, in early 2010, a network in

---

<sup>2</sup> Cf. Die Welt.de: <https://www.welt.de/politik/deutschland/article218298328/GroKo-Beschluss-Geheimdienste-duerfen-nun-WhatsApp-Chats-mitlesen.html>, last Accessed 26.01.2021

<sup>3</sup> Cf. Der Standard: <https://www.derstandard.at/story/2000118906392/wie-deutsche-ermittler-bei-whatsapp-mitlesen-koennen>, last Accessed 26.01.2021

<sup>4</sup> Cf. Die Presse: <https://www.diepresse.com/687375/kriminalitat-virtuelle-spielewelten-als-tatorte-fur-verbrecher>, last Accessed 26.01.2021

<sup>5</sup> Cf CBSN: <https://www.cbsnews.com/video/terrorists-use-video-games-to-communicate-undetected/>, last Accessed 26.01.2021



which sub-tokens (metatokens) can be generated for a specific application was developed. The Blockchain systems NXT and/or Ardor<sup>6</sup> and Ethereum<sup>7</sup> are particularly noteworthy in this context from a historical as well as current perspective.

Blockchain, in the sense of cryptocurrencies are found in the cybercrime literature and discussion primarily in the context of terrorist financing, the movement of funds and money laundering, but not in the scope of peer2peer communication. This also seems to make sense at first glance, as blockchain systems are characterized by information being stored forever on the one hand, and on the other hand, more and more countries have implemented strict know-your-customer regulations when it comes to converting the proceeds from the sale of blockchain tokens into FIAT currency (a currency established as money, often by government regulation). These steps are, of course, to be welcomed as it helps legitimize cryptocurrencies in regards to transnational trade.

However, for the scope of this paper, the authors would like to take a closer look at testnet systems of blockchains, more specifically how these can be easily set up by evildoers as completely private systems, operated off the grid of authorities. Also, it is important to note how the mainnet systems of public blockchains, especially those on which meta-tokens can be created, can be used to communicate encrypted and completely unnoticed. In addition, the authors would like to discuss how new technologies such as pruning, where the private information, i.e. the encrypted text message, can no longer be stored by the blockchain after a certain self-defined block height, increases the possibility of illegal activities via such testnet systems of different blockchains. This will in turn strongly address the issue of data ownership mentioned previously, because especially in the case of private networks, which in turn rely on technologies such as pruning within their network, the intrinsically positive aspect of data security and ownership could result in major barriers to investigation from the authorities' point of view.

## **2. Related research**

In terms of communication and terrorism, most of the literature focusses on terrorists engaging in communication with the outside world. Matusitz (2013) described this from the different viewpoints, while Mahmood and Jetter (2020) analyzed the role of communication technology from 1970 to 2014, finding that online communication via internet is an emerging tool to spread the word to their followers. A similar aspect, often covered in literature is the role of media, reporting about terrorism. For instance, this is addressed by Archetti (2013) who focuses on the aspect of news agencies producing headlines that might produce many "reads", but whose content is not based on facts. Cahyan et. al (2017) research highlighted the importance of mobile device forensics in investigations involving the use of cloud storage services and communication apps along with the necessity and potential utility of the integrated incident handling and digital forensics models to investigate and reconstruct terrorist incidents. Their research included the investigation of three popular cloud apps (Google Drive, Dropbox and OneDrive), five communication apps (Messenger, WhatsApp, Telegram, Skype and Viber), and two email apps (GMail and Microsoft Outlook). However, blockchain as communication tool is an original and innovative idea that is not found in literature. This is probably explained by the fact that blockchain, as already mentioned, was originally designed to store data forever and immutably. However, the authors take a special look at testnet systems and new technologies such as pruning, where attached records are also deleted after a self-defined block height. Therefore, the intended research described in this WIP paper is an absolute novelty.

## **3. Planned research goals and methods used**

Therefore, the aims of the upcoming research are:

- 1. To analyse the various Blockchain testnet systems in detail from different points of view: How difficult is it to install your own full node, how difficult is it to create a wallet, how can your own meta tokens be generated, how can encrypted messages be attached to the native tokens of the system or the meta tokens created by the users when transferring them, which encryptions are used in the process and which are used for the wallet-address of the user. Is pruning as technology implemented in the network? And above all, how users get test tokens of the native token system and what information is revealed about the user in the process? This analysis is done by installing and testing the common Blockchain testnet systems on the

---

<sup>6</sup> Jelurida: <https://www.jelurida.com>, last Accessed 26.01.2021

<sup>7</sup> More on Ethereum: <https://ethereum.org>, last Accessed 26.01.2021

one hand and by literature review including the whitepapers related to the respective networks on the other hand. In addition, resources such as Git-Hub will be consulted in order to evaluate specific functionalities.

- 2. To interview experts from different fields and perform a content analysis of the answers based on the results of (1).
- 3. To develop guidelines with suggestions for the blockchain community, as well as the authorities and cybercrime researchers, as to what measures should be taken in the future, for example in the form of a know your customer (KYC) light process when distributing blockchain testnet tokens to the users/developers of the corresponding testnet systems, including zero-knowledge proof systems for authentication and shared-key options – for example for multi-signature from different authorities – to access the necessary data only in a well-founded suspicious case.

These steps should always be crafted under the premise that blockchain systems have the potential to contribute very positively to the world of data security. As such, future regulations should therefore not interfere with the development of new applications via the testnet systems. Nevertheless, they should provide a necessary hurdle to make the use of these systems less attractive to cyberterrorists.

#### **4. Conclusion**

With this upcoming article, the authors want to contribute to the field of cybercrime and initiate a new discussion on this topic. The research will take place in the second half of 2021 and the final paper will be published and presented in 2022.

#### **References**

- Archetti C. (2013) Terrorism, Communication, and the Media. In: Understanding Terrorism in the Age of Global Media. Palgrave Macmillan, London. [https://doi.org/10.1057/9781137291387\\_3](https://doi.org/10.1057/9781137291387_3)
- Cahyani, N. Ab, R.; Glisson, W.; Choo, K. (2017). The Role of Mobile Forensics in Terrorism Investigations Involving the Use of Cloud Storage Service and Communication Apps. *Mobile Networks and Applications*. 22. 10.1007/s11036-016-0791-8.
- Cahyani, Niken Dwi; Wahyu; Rahman, Nurul Hidayah; Ab; Glisson, William Bradley; Choo, Kim-kwang Raymond (2017) .*Mobile Networks and Applications*; New York Bd. 22, Aug. 2, (Apr 2017): 240-254. DOI: 10.1007/s11036-016-0791-8
- EFAIL: EFAIL describes vulnerabilities in the end-to-end encryption technologies OpenPGP and S/MIME that leak the plaintext of encrypted emails. (2018) Retrieved from <https://efail.de/#paper>, last Accessed 26.01.2021
- Grech, A. and Camilleri A. (2017) Blockchain in Education. <https://doi.org/10.2760/60649>; Accessed: January, 2020
- GSMA Mobile for Development Foundation, Inc. (2013) Retrieved from [https://www.gsma.com/publicpolicy/wp-content/uploads/2013/11/GSMA\\_White-Paper\\_Mandatory-Registration-of-Prepaid-SIM-Users\\_32pgWEBv3.pdf](https://www.gsma.com/publicpolicy/wp-content/uploads/2013/11/GSMA_White-Paper_Mandatory-Registration-of-Prepaid-SIM-Users_32pgWEBv3.pdf), last Accessed 26.01.2021
- Mahmood, R., & Jetter, M. (2020). Communications Technology and Terrorism. *Journal of Conflict Resolution*, 64(1), 127–166. <https://doi.org/10.1177/0022002719843989>
- Matusitz, J. (2013) *Terrorism & Communication, a Critical Introduction*, Sage, Los Angeles
- Nofer, M., Gomber, P., Hinz, O. and D. Schiereck (2017), *Blockchain*, *Bus. Inf. Syst. Eng.*, vol. 59, no. 3, pp. 183–187, Mar. 2017. DOI 10.1007/s12599-017-0467-3
- Satoshi Nakamoto (2008) Bitcoin: A Peer-to-Peer Electronic Cash System, in Whitepaper online available <https://bitcoin.org/bitcoin.pdf> (Satoshi Nakamoto is a pseudonym, it is not known to the general public who is behind this name.); Accessed: January, 2020

# Ethics of Cybersecurity in Digital Healthcare and Well-Being of Elderly at Home

Jyri Rajamäki

Laurea University of Applied Sciences, Espoo, Finland

[jyri.rajamaki@laurea.fi](mailto:jyri.rajamaki@laurea.fi)

DOI: 10.34190/EWS.21.009

**Abstract:** The SHAPES Horizon 2020 project supports the well-being of the elderly at home. The growing complexity of the digital ecosystem in combination with increasing global risks involves various ethical issues associated with cybersecurity. An important dilemma is that overemphasising cybersecurity may violate fundamental values such as equality and fairness, but on the other hand, neglecting cybersecurity could undermine citizens' trust and confidence in the digital infrastructure, policymakers and state authorities. One example of ethical issues concerning health and well-being is that if a medical implant producer protects the data transfer between implant and receiver server utilising suitable cryptology, this significantly increases the energy consumption of the implant and frequently requires more surgeries for battery exchange. The object of this work in progress paper is to help to provide necessary tools and guidelines to health and well-being service developers in the SHAPES project for their ethical consideration of cybersecurity actions. This paper examines different views and approaches to the ethics of cybersecurity in healthcare and finds the most relevant and puzzling issues for the SHAPES project. The paper investigates the ethical issues, for example, applying the approach of principlism based on four principles of biomedical ethics (respect for autonomy, nonmaleficence, beneficence and justice), and ethics of care. The important aims of the employment of information and communication technology in healthcare are efficiency and quality of services, the privacy of information and confidentiality of communication, the usability of services, and safety. Four important value clusters in cybersecurity are security, privacy, fairness, and accountability. From these four different ethical aspects (biomedical ethics, ethics of care, core value clusters in cybersecurity, and technical aims), this paper proposes a new conceptual model for a system approach to analyse the ethical matters, which are related to cybersecurity in digital healthcare and well-being.

**Keywords:** ethics, cybersecurity, digital healthcare, SHAPES project, healthy ageing, well-being

---

## 1. Introduction

Digital transformation and ecosystem thinking steer the Smart and Healthy Ageing through People Engaging in Supportive Systems (SHAPES) Horizon 2020 project that supports the well-being of the elderly at home. From an ethics point of view, SHAPES is a diverse solution and ethical requirements and their implementation are essential for the sustainability of SHAPES. The implementation of ethical requirements has an impact not only on technical solutions and services but also on the organisational arrangements of SHAPES. Alongside user requirements, ethical requirements are particularly important when developing solutions linked to fundamental rights, and when the target group is older persons (Sarlio-Siintola, 2020).

The paper is structured as follows. After the introduction, the literature review investigates four different ethical aspects related to cybersecurity in digital healthcare and well-being: biomedical ethics, ethics of care, core value clusters in cybersecurity, and technical aims of Information and Communication Technology (ICT) systems in healthcare. The third section proposes a new conceptual model for a systematic analysis of relations between these different ethical aspects. The last section discusses future work.

## 2. Ethical frameworks related to digital healthcare and well-being

### 2.1 Core values in cybersecurity

According to van de Poel (2020), four important value clusters exist that should be considered when deciding on cybersecurity measures. The first one 'security' is a combination of more specific values, such as individual security, national resilience and information security. These values protect humans and other valuable entities from all kinds of harm and respond to morally problematic situations in which harm is done, ranging from data breaches and loss of data integrity to cybercrime and cyberwarfare (van de Poel, 2020).

The second value cluster 'privacy' contains such values as privacy, moral autonomy, human dignity, identity, personhood, liberty, anonymity and confidentiality. According to van de Poel (2020), these values correspond to the following norms: "we should treat others with dignity, we should respect people's moral autonomy, we should not store or share personal data without people's informed consent, and we should not use people (or

data about them) as a means to an end.” Moral problems with these values include the secret collection of large amounts of personal data for cybersecurity purposes or the unauthorised transfer of personal data to a third party (van de Poel, 2020).

The third cluster ‘fairness’ consists of values such as justice, fairness, equality, accessibility, freedom from bias, non-discrimination, democracy and the protection of civil liberties. These values respond to the fact that cybersecurity threats, or measures to avoid them, do not affect everyone equally being sometimes morally unfair. Another moral problem is that cybersecurity threats, or measures to increase cybersecurity, may undermine democracy, civil rights and liberties. Moral reasons that correspond to these values are that people should be treated fairly and equally, and democratic and civil rights should be upheld (van de Poel, 2020).

The fourth cluster ‘accountability’ includes values such as transparency, openness and explainability. If governments take cybersecurity measures that harm citizens and require the weighing of a range of conflicting substantive values such as security, privacy and fairness, then accountability, as a more procedural value, is particularly relevant (van de Poel, 2020).

In addition to the four value clusters, some domain-specific ethical principles and values are different from domain to domain, and technical aims can be different even from application to application. They are connected to a range of instrumental or technical values related to the proper functioning of applications such as efficiency, ease of use, understandability, data availability, reliability, compatibility and connectivity. However, technical values are morally relevant as they are instrumental for achieving moral values (van de Poel, 2020).

## 2.2 Biomedical ethics

Biomedical ethics is an interdisciplinary, contemporary ethical approach based on Beauchamp and Childress’s (2009) four main principles: justice, beneficence, non-maleficence, and autonomy. It serves as a paradigm that assists healthcare professionals and public policymakers to identify and respond to moral dilemmas in biomedical and healthcare research and encompasses different types of moral norms: moral ideals, virtues, rules, and principles. Principles are considered general norms, and they leave considerable space for judgement in several cases. Principles do not function as ‘precise action guides’ that would inform us in every single circumstance on how to act the same way as detailed judgements and rules would guide. The principles are rather abstract, and they do not form a general moral theory but a framework to identify and reflect on moral problems (Sarlio-Siintola, 2020).

## 2.3 Ethics of care

The care sector applies ‘Ethics of care’ based on Gilligan’s (1982) ideas that there are two different types of moralities: the ethic of justice and the ethic of care. Gilligan (1982) explains, “the ethic of care is centred on maintaining relationships through responding to needs of others and avoiding hurt”. Care ethics see moral problems arising from ruptures or tensions in relationships. Within care reasoning, moral problems are solved by considering the unique characteristics of situations and persons, more than applying a hierarchy of rights or rules; the latter would be more typical of a justice ethics approach. The nursing field greets Gilligan’s theory with enthusiasm, as it has “theoretically captured the essence of caring embedded in patient-nurse relationships and explained the ethical difficulties nurses encountered in medically dominated healthcare contexts” (Juujärvi, et al., 2019). It is a promising approach to strengthen the voices of nurses in ethical discussions, in which justice-based theories traditionally dominate.

Table 1 presents the main characteristics of care ethics in the SHAPES context.

**Table 1:** Main characteristics of care ethics (Sarlio-Siintola, 2020)

<i>Perspectives</i>	<i>In the SHAPES context, especially</i>
Empathy	Showing empathy might need new forms when acting on digital platforms: e.g., a smile, touch and eye contact might not work as in traditional face-to-face encounters – this applies to caregivers, researchers and older persons.
Relationships	Building and maintaining relationships might mean learning new methods and forms when acting on digital platforms. Building and maintaining relationships also means an understanding of, e.g., psychology, sociology and spirituality of human beings.

Perspectives	In the SHAPES context, especially
Uniqueness of the case	In hectic working life, it might not always be easy to provide care, as the case is unique and not just one of a dozen similar-looking ones.

## 2.4 Desiderata of ICT in health and the instrumental role of cybersecurity

Four main functions of ICT systems in healthcare are: improving the quality and efficiency of services, protecting confidentiality, enhancing usability, and protecting patients’ safety. Weber and Kleine (2020, 143-145) summarizes these functions as follows:

- 1. “One of the main purposes of ICT systems in healthcare is the administration of information to increase the *efficiency* of the healthcare system and to reduce its costs. Improvements in healthcare in *qualitative* terms refer, for instance, to new services that provide treatment or processes with better health-related outcomes. Big Data, the collection and sharing of as much health-related data as possible might be used to establish new insights regarding diseases and possible treatments.”
- 2. “Using ICT to process patient data creates a moral challenge in terms of quality on the one hand and *privacy* and confidentiality on the other hand—yet both are important aims in healthcare. In particular, privacy is often seen as a prerequisite of patients’ autonomy”...“Privacy and confidentiality are also foundations of trust among patients on the one hand and healthcare professionals on the other.”
- 3. Roman, et al. (2017) define *usability* as the degree of effectiveness, efficiency, and satisfaction with which users of a system can realize their intended task. Concerning health, users include patients, medical staff and/or administrators, which have different degrees of ICT competences, depending on personal attitudes and socio-demographic variables (Weber & Kleine, 2020).
- 4. “*Safety* can be defined as the reduction of health-threatening risks. Safety, quality, efficiency and usability are interrelated, but they do not align, because safety measures might reduce the efficiency and usability of services and therefore quality.”

The instrumental role of cybersecurity in healthcare is to protect against three types of threats based on the target of the attack: threats against information, information systems and medical devices (Loi, et al., 2019).

## 3. Conceptual model for systematic analysis of the ethics of cybersecurity in healthcare

Figure 1 proposes a new conceptual model for a systematic relation analysis of ethical matters related to cybersecurity in digital healthcare and well-being. The systematic mapping of the relations between the four different ethical aspects (biomedical ethics [n=4], care ethics [n=3], core value clusters in cybersecurity [n=4] and technical aims [n=4]) generates 84 value pairs.

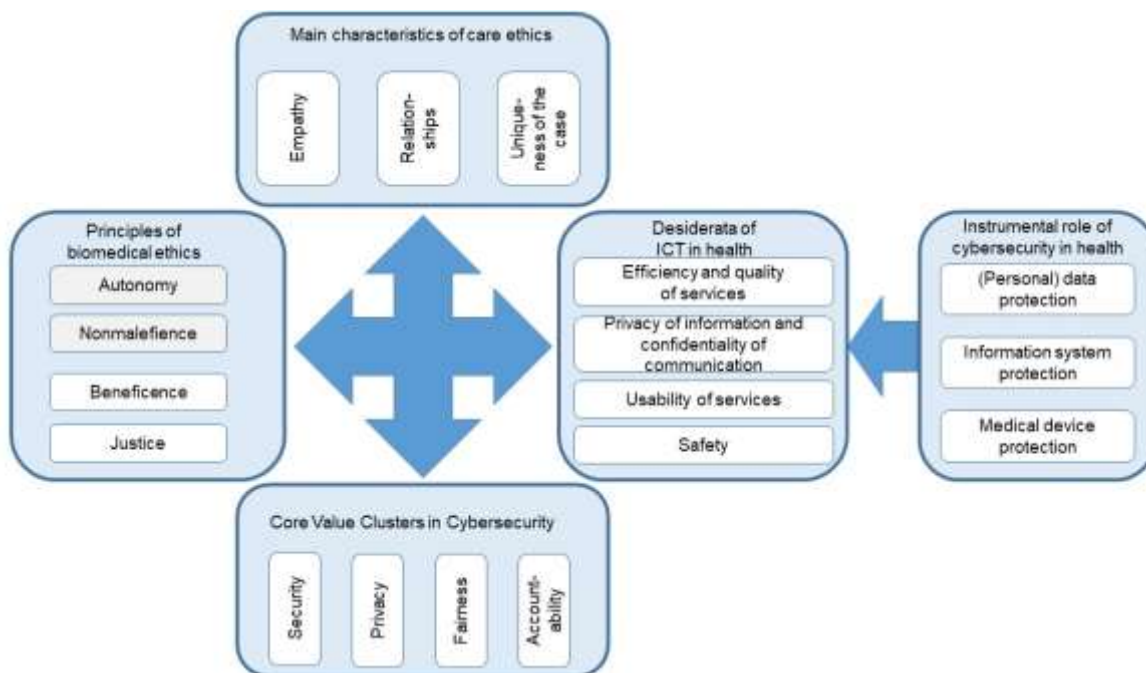


Figure 1: Conceptual model for analysing ethical aspects of cybersecurity in healthcare

#### 4. Discussion

Ethics is crucial in healthcare and new eHealth services make ethical questions even more pressing and raises new ones, such as ethics of cybersecurity in healthcare (Weber & Kleine, 2020). Loi et al. (2019) have investigated the relation between ICT desiderata and the four principles of medical ethics and mapped trade-offs between the goals of cybersecurity into conflicts between the four principles of medical ethics. A similar analysis is needed from the relations between (1) biomedical ethics vs. ethics of care, (2) biomedical ethics vs. core values in cybersecurity, (3) ethics of care vs. technical aims, (4) ethics of care vs. core values in cybersecurity, and (5) technical aims vs. core values in cybersecurity.

#### Acknowledgements

This work was supported by the SHAPES project, which has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement no. 857159.

#### References

- Beauchamp, T. & Childress, J. 2009. *Principles of biomedical ethics*. New York: Oxford University.
- Gilligan, C. 1982. *In a Different Voice: Psychological Theory and Women's Development*. Cambridge: Harvard University Press.
- Juujärvi, S., Ronkainen, K. & Silvennoinen, P. 2019. The ethics of care and justice in primary nursing of older patients. *Clinical Ethics*, 14(4), 187–194.
- Loi, M., Christen, M., Kleine, N. & Webe, K. 2019. Cybersecurity in health—disentangling value tensions. *Journal of Information, Communication and Ethics in Society*, 17(2), 229-245.
- Roman, L., Ancker, J., Johnson, S. & Senathirajah, Y., 2017. Navigation in the electronic health record: A review of the safety and usability literature. *Journal of Biomedical Informatics*, Volume 67, pp. 69-79.
- Sarlio-Siintola, Sari (ed.). 2020. *SHAPES Ethical Framework D8.4*. [online] Available at: <https://shapes2020.eu/wp-content/uploads/2020/11/D8.4-SHAPES-Ethical-Framework.pdf>
- van de Poel, I. 2020. Core Values and Value Conflicts. In: *M. Christen et al. (eds.), The Ethics of Cybersecurity*. Cham: Springer, pp. 45-72.
- Weber, K., & Kleine, N. 2020. Cybersecurity in Health Care. In: *M. Christen et al. (eds.), The Ethics of Cybersecurity*. Cham: Springer, pp. 139-156.

# ECHO Federated Cyber Range as a Tool for Validating SHAPES Services

Jyri Rajamäki and Harri Ruoslahti

Laurea University of Applied Sciences, Espoo, Finland

[jyri.rajamaki@laurea.fi](mailto:jyri.rajamaki@laurea.fi)

[harri.ruoslahti@laurea.fi](mailto:harri.ruoslahti@laurea.fi)

DOI: 10.34190/EWS.21.076

**Abstract:** ECHO is a cybersecurity pilot project under the H2020 Program. The ECHO Federated Cyber Range (E-FCR) provides enabling technology supporting ECHO Network operations, ensuring a safe and reliable multi-sector simulation environment in which to ensure viable delivery of identified technology roadmaps, as well as, hands-on cyber-skills development involving realistic sector specific or multi-sector simulations. A cyber range leverages cloud technologies to provide a virtualized environment in which realistic cyber scenarios can be instantiated. The eHealth platform by project SHAPES will rely on services and products provided by vendors. Operability and usability of the platform requires reliable, uninterrupted and well-managed actions from the systems utilized to run services of the eHealth platform. The SHAPES platform operates in the cyber domain and the taxonomy of cyber-risks vary from actions of people due lack of cybersecurity awareness to technology failures. Moreover, threats from malicious external sources might exploit vulnerabilities of SHAPES assets and therefore cause damage. Predefined security validation procedures facilitate to create a baseline for services and their desired level of security. This work-in-progress paper explores how to apply E-FCR during eHealth-services validation processes. The paper profits two Horizon-2020 projects: The ECHO cybersecurity project demonstrating how to utilize E-FCR in the healthcare domain; and the SHAPES healthcare project that needs a cybersecurity validation processes for services incorporated into the SHAPES platform.

**Keywords:** ECHO project, SHAPES project, federated cyber range, security validation

---

## 1. Introduction

Data breaches in the healthcare sector are occurring at unprecedented rates, and the causes of these breaches have not delineated very well (McLeod & Dolezel, 2018), so protecting the cyber environment of healthcare systems and operators is important in protecting society and its critical infrastructure.

Project *Smart and Healthy Ageing through People Engaging in Supportive Systems (SHAPES)* gathers stakeholders from across Europe to create, deploy and pilot at large-scale an EU-standardized open platform integrating a broad range of technological, organizational, clinical, educational and societal solutions. The aim is to enable ageing Europeans to remain healthy, active and productive, while maintaining a high quality of life and sense of wellbeing for the longest time possible (European Commission, 2019). *European network of Cybersecurity centres and competence Hub for innovation and Operations (ECHO)* is one of four European pilot projects, which together aim to establish and operate a European network of cybersecurity excellence. During its four-year life span, ECHO will develop and deliver an organized and coordinated, effective and efficient multi-sector collaboration based approach that helps strengthen the proactive cyber defences of the European Union (Pappalardo et al., 2020).

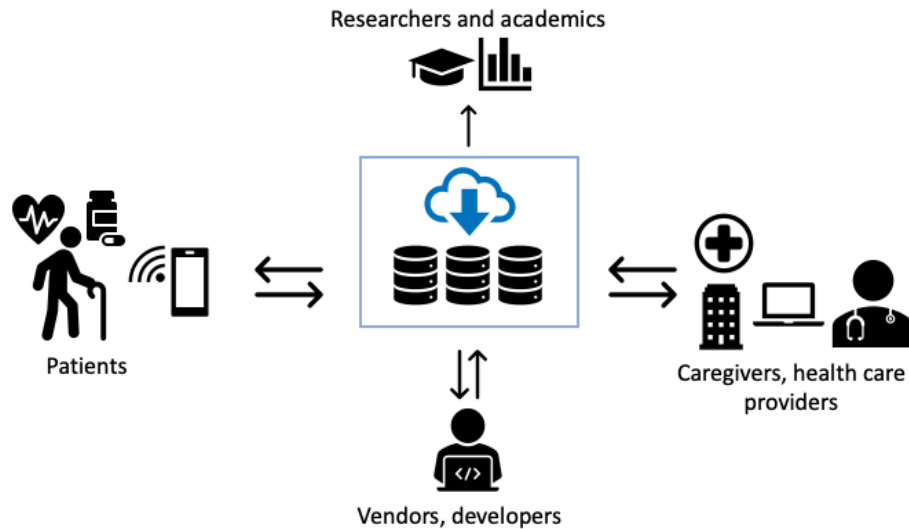
In January 2021, the ECHO project started working on the Demonstration Cases that are key to validate ECHO technology roadmaps and their combination. The SHAPES project needs a cybersecurity validation processes for services incorporated into the SHAPES platform. This work-in-progress paper explores possibilities to apply ECHO's Federated Cyber Range (E-FCR) to validate eHealth-services. This paper is organized as follows: Section 2 deals with cybersecurity of eHealth platforms. Section 3 outlines security validation requirements for eHealth services. Section 4 presents ECHO's work in the healthcare sector and E-FCR. Section 5 concludes the paper and suggests possibilities for future work.

## 2. eHealth platform

ENISA, the European Union Agency for Cybersecurity (2020) has listed various threats to hospitals including system, device, software or network component failures, human errors, and malicious actions such as denial of service or web application attacks. Supply chain failures, cloud service provider failures especially, are an important threat source (ENISA, 2020). The SHAPES eHealth platform combines in-formation technology and communication components to support the well-being of elderly people, mainly technologies and information

or primarily clinical and medical purposes, while academic study and decision-making can use secondary information.

ENISA (2015) identifies critical platform assets: Health information systems; Databases; Authentication server; Laboratory and radiology information systems; Electronic health record components and service; ePrescription services.



**Figure 1:** eHealth ecosystem

Figure 1 above presents a simplified illustration of the eHealth platform. Health data is centralized to the SHAPES Secure Cloud that would act as a hub between ecosystem stakeholders, store patient data, and allow personalised services for patients. Aggregated big data from patients would be used secondarily after modifications to ensure anonymous origin of data/information.

The eHealth ecosystem includes the critical assets (ENISA, 2015): healthcare providers have components of network devices, patient and caregiver use web applications to access information stored in data-bases and both authenticate themselves to services through an identity management system. Patients have devices to save daily results of exercise and blood-pressure for later evaluation.

Baselines for information security policies e.g. acceptable use policy for services and general security awareness requirements (ENISA, 2020) should be addressed by SHAPES so that ecosystem stakeholders understand the importance of safeguarding sensitive information and what risks are related to mishandling information. Figure 2 presents how PCI Security Standards Council (2014) relates risk levels, roles with depth of security awareness training, emphasising the importance of awareness training as part of overall risk management activities.



**Figure 2:** Security awareness and level of risk (PCI Security Standards Council, 2014)



### 3. Requirements for validating eHealth services

Information and data protection requirements for social and healthcare procurement by Finnish National Cyber Security Centre (NCSC-FI) are presented in themes under headings. These requirements are directional but not definite for every procurement (Traficom, 2019). The original file with requirements can be accessed from NCSC-FI web page. The following requirements may help mitigate risks against SHAPES assets. ENISA (2020) has similar guidelines for procurements in hospitals. ENISA guidelines have been separated to three phases: plan, source and manage in addition to general good IT practices. All phases encourage hospitals to embrace good practices for cybersecurity. Table 1 presents similarities between these two publications:

**Table 1:** Similarities in NCSC-FI requirements and ENISA guidelines

Phases in ENISA's guidelines	ENISA	NCSC-FI
General practices	Involve IT department in procurement Vulnerability management Policy for hardware and software Secure wireless communication Establish testing policies Establish Business Continuity Plans Consider interoperability issues Allow auditing and logging Use encryption	Wireless systems Incident support Secure architecture Logging Certificate and key management Interface security On-premises installations Inform personnel
Plan phase	Conduct risk assessment Plan requirements in advance Identify threats Segregate network Establish eligibility criteria for suppliers Create dedicated RfP for cloud	Risk management Segmentation & data flows Data protection & safekeeping 3 <sup>rd</sup> party software Security control
Source phase	Require certification Conduct DPIA Address legacy systems Provide cybersecurity training Develop IRP Involve supplier in incident management Organise maintenance operations Secure remote access Require patching	Security patch management System administration Personal data protection On-premises installations Incident support Security control System privacy Personnel security contracts
Manage phase	Raise cybersecurity awareness Perform asset inventory and configuration Dedicated access control mechanisms Schedule penetration testing frequently or after modifications in the architecture/system	Security testing System & security monitoring User accounts & authentication User rights & session management Change management 3 <sup>rd</sup> party software Inform personnel Organisational and personnel changes System and information disposal

### 4. ECHO Federated Cyber Range (E-FCR)

ECHO develops cybersecurity technology roadmaps based on analysis of current and emerging cybersecurity challenges and associated technologies. These roadmaps may serve as foundations for novel industrial capabilities, and in developing new and innovative technologies to address the cybersecurity challenges of Europe. Research and development of early ECHO prototypes will target high-priority opportunities specified in the six ECHO roadmaps, one of which is the E-FCR (Kirkov et al., 2020).

E-FCR combines capabilities of several independent interconnected cyber-ranges, and can help simulate inter-sector scenarios that include complex realities and inter-sector dependencies (Kirkov et al., 2020). The healthcare sector, which offers life-critical services, is increasingly adopting new technologies, which while they intend to improve treatment and care, they may also be vulnerable to cyber threats (McLeod & Dolezel, 2018). In the health sector cyber-incidents not only threaten the security of medical systems and information, but

patients' lives. New medical technology adds value to healthcare only when it is (cyber) secure (Pappalardo et al., 2020).

ECHO recommends reducing complexity in healthcare systems, raising awareness, and specifying cybersecurity skills and training curricula for all levels of healthcare staff. Hospital and organizations should budget for increased and cybersecurity risk assessment and management (Pappalardo et al., 2020).

The E-FCR user experience is based on gamification; "the use of game design elements characteristic for games (rather than play or playfulness) in non-game contexts" (Kirkov et al., 2020, p.29). Gamification thus, provides experiences of social engagement, which in turn feeds into how we look for competition between individuals or teams (Dankbaar et al., 2017; Skerlavaj, Dimovski, & Desouza, 2010). Team projects, group learning opportunities, competition, cooperation, collaboration, and freedom of choice promote organizational learning (Ruoslahti & Trent, 2020) and co-creation (Ruoslahti, 2018).

Cyber-threat lists may be based on previous incidents, threat assessments from e.g. industry/standardization/governmental bodies, and information from social media and other available sources and repositories can provide augmenting information that helps rapidly draw conclusions to build timely cyber situational awareness (Pöyhönen et al., 2020). "The E-FCR platform should search for current technological threats and vulnerabilities on the internet by collecting information of the high priority vulnerabilities. Moreover, it should regularly check the latest available products and capabilities of all Cyber Range providers in the marketplace" (Kirkov et al., 2020, p.35).

## **5. Conclusions and suggestions**

ICT is becoming more and more pervasive in the healthcare sector including computerized systems for automation of diagnostic and collection of patient data. Sensors and medical devices with IP addresses are connected to the Internet (IoT). Multidisciplinary teams interact with patient and share sensitive data also through personal devices. Predefined security requirements facilitate the creation of baselines for services and achieving desired levels of security. NCSC-FI requirements and ENISA guidelines for procurements encourage healthcare organisations set requirements for vendors. These requirements and recommendations are well structured and easy to follow but require revision and professional assessment for each procurement as nor NCSC-FI or ENISA guidelines are definite for all organisations. It is an important to verify that vendors truly meet the security requirements throughout the service or product lifecycle.

Examples of use-cases to be implemented by the ECHO project are 1) attacks against complex medical systems (blood analysis laboratory), and 2) attacks against connected physical medical devices. ECHO will create at least two sector-specific cyber-range to support healthcare-sector demonstration cases: 1) HC Cyber-Range (blood analysis laboratory) which is already ready in RHEA premises, 2) the other cyber range will leverage several medical devices. Taking into account lessons from SHAPES should enrich the ECHO demonstration cases.

## **Acknowledgements**

This work was supported by the ECHO project, which has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement no 830943.

## **References**

- Dankbaar, M. et al. 2017. Comparative effectiveness of a serious game and an e-module to support patient safety knowledge and awareness. BMC Medical Education. <https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-016-0836-5>
- ENISA. 2020. Procurement Guidelines for Cybersecurity in Hospitals. European Union Agency for Cybersecurity. <https://www.enisa.europa.eu/publications/good-practices-for-the-security-of-healthcare-services>
- ENISA. 2015. Security and Resilience in eHealth Infrastructure and Services. European Union Agency for Cybersecurity. <https://www.enisa.europa.eu/publications/security-and-resilience-in-ehealth-infrastructures-and-services>
- European Commission. 2019. "Smart and healthy ageing through people engaging in supportive systems," [Online]. <https://cordis.europa.eu/project/id/857159>
- Kirkov, P. et al. 2020. D4.3 Inter-Sector Cybersecurity Technology Roadmap. <https://echonetwork.eu/deliverables/>
- McLeod, A., & Dolezel, D. 2018. Understanding Healthcare Data Breaches: Crafting Security Profiles.
- Pappalardo, M. et al. 2020. D2.2 ECHO Multi-Sector Assessment Framework. <https://echonetwork.eu/deliverables/>
- PCI Security Standards Council. 2014. Information Supplement: Best Practices for Implementing a Security Awareness Program.

***Jyri Rajamäki and Harri Ruoslahti***

[https://www.pcisecuritystandards.org/documents/PCI\\_DSS\\_V1.0\\_Best\\_Practices\\_for\\_Implementing\\_Security\\_Awareness\\_Program.pdf](https://www.pcisecuritystandards.org/documents/PCI_DSS_V1.0_Best_Practices_for_Implementing_Security_Awareness_Program.pdf)

Pöyhönen, J., Rajamäki, J., Ruoslahti, H., & Lehto, M. 2020. Cyber Situational Awareness in Critical Infrastructure Protection. *Annals of Disaster Risk Sciences*, 3(1).

Ruoslahti, H. & Trent, A. 2020. Organizational Learning in the Academic Literature – Systematic Literature Review. *Information & Security: An International Journal* 46, no. 1 (2020): pp. 65-78.

Ruoslahti, H. 2018. Co-creation of Knowledge for Innovation Requires Multi-Stakeholder Public Relations, in Sarah Bowman, Adrian Crookes, Stefania Romenti, Øyvind Ihlen (ed.) *Public Relations and the Power of Creativity (Advances in Public Relations and Communication Management, Volume (3) Emerald Publishing Limited*, pp.115 – 133.

Traficom. 2019. Information security and data protection requirements for social welfare and healthcare procurements. <https://www.kyberturvallisuuskeskus.fi/en/ncsc-news/instructions-and-guides/information-security-and-data-protection-requirements-social>