Journal of Information Warfare

Volume 20, Issue 1
Winter 2021

Contents

Journal of Information Warfare

© Copyright 2020

Published by
Peregrine Technical Solutions, LLC
Yorktown, Virginia, USA

Print Version
ISSN 1445-3312

Online Version
ISSN 1445-3347

From the Editor

Fall 2020

This issue celebrates the best papers from the International Conference on Cyber Warfare and Security (ICCWS) held last spring at Old Dominion University (ODU) in Norfolk, VA. As the dates for the conference neared (9-13 March 2020), the United States (and the world) was rapidly shutting down due to COVID-19. Even with 50+ delegates from all over the world already onsite, it was touch and go up to almost the very last moment as to whether the event was going to happen. Luckily, we were able to proceed with the conference, which might go down as one of the last academic conferences conducted in person in the U.S. before the lockdown. What is certain is that the conference was a great event, as evidenced by the papers here that represent the best of the best as rated by the conference stream chairs.

Since those five days in early March, the citizens of the world have struggled to confront a global pandemic the likes of which have not been seen in a century, and the end of which (we are now being cautioned) may be nowhere in sight. That struggle has been and continues to be instructive. Some countries have done quite well in their handling of what has been called the first wave of the outbreak … and some have not. This difference alone is remarkable. Consider how countries such as New Zealand and South Korea have managed to develop and carry out mitigating protocols that have kept their casualty rates among the lowest in the world.

From an information-warfare perspective, it is also remarkable how the pandemic has been politicized and weaponized by different factions, groups, and countries. In the United States, the presidential election has become a referendum on the handling of this crisis and a demonstration of the deep divide that exists between political parties. Elsewhere, it has been reported that a number of countries, including Iran, are vastly undercounting and underreporting their numbers of COVID deaths. And far ahead of the rest of the world, Russia has, it says, developed a “new vaccine” but offers little medical evidence of or information about its efficacy.

Together, what these vignettes demonstrate is how information can be used for good … and bad. Across the globe, in response to this pandemic, we’ve witnessed the fruits of cooperation and transparency, as well as the consequences of division, disinformation, and deception. Life-saving and life-threatening information has been weaponized to advance agendas and achieve political outcomes.

Meanwhile, winter is coming.

Stay well,

Dr. Leigh Armistead, CISSP
Chief Editor, Journal of Information Warfare
leigh.armistead@goldbelt.com

Authors

Prof. Virginia Greiman is an internationally recognized scholar and expert in the fields of national cyber security and cyber law and regulation. She serves as Assistant Professor at Boston University Metropolitan College and is a member of the Boston University Law Faculty. She has also held academic appointments at Harvard University Law School and Harvard Kennedy School of Government. Her teaching and research focus on megaproject strategies and governance, cyber law and international law, national security strategies, cyber warfare and surveillance, global cybercrime and enforcement, privacy law and big data, and corporate innovation and competitiveness. She served in the United States Department of Justice and as an international legal consultant for the U.S. Department of State in Eastern and Central Europe. She has served as an advisor to numerous international and national organizations including the United States Air Force Institute of Technology Center for Cyberspace Research, the United States Agency for International Development, the National Aeronautics and Space Administration (NASA), the World Bank, and the United Nations Economic and Social Council (ECOSOC).

Professor Greiman has held executive and advisory positions with several of the world's largest megaprojects in the United States, Europe, Africa, and Southeast Asia including Boston's $15 billion-dollar Big Dig Project, London's Crossrail Project, India's Megaproject Initiatives and Smart Cities Program, Taiwan and Southeast Asia National Science Parks, the U.S. Nuclear Power Industry and Development in the South China Sea.

Dr. John S. Hurley serves as a Professor in the College of Information and Cyberspace at the National Defense University. Hurley has over 35 years' experience in the area of information and computing technologies. He served as Senior Manager, Distributed Computing in the Networked Systems Division, for the Boeing Company, Bellevue, WA. Dr. Hurley was Professor of Electrical Engineering and Director of three research centers (Scalable and Embedded Applications Center, Materials Processing Assessment and Characterization Center, and Avalon Scalable Embedded Computing Center) and the Co-Director, Army Center of Excellence in Electronic Sensors and Combat at Clark Atlanta University, in Atlanta, GA. He is a 2015 Seminar XXl Fellow.

Gazmend Huskaj is a doctoral student in Cyberspace Operations at the Swedish Defence University. Previously, he was Director of Intelligence in the Swedish Armed Forces on cyber-related issues. Prior to that, he was Head of the United Nation's Intelligence Cell in a mission area for several years. He is a veteran, with more than five years of duty in conflict and post-conflict areas including two tours to the Balkans and one in Central Asia. He is a graduate from Harvard Kennedy School in Cybersecurity: The Intersection of Policy and Technology, and Geneva Centre for Security Policy European Training Course (ETC). In 2014, he was awarded the best idea answering to EEAS Deputy Secretary General's thread on the EU as a security provider. Gazmend holds a two-year-master's (MSc) in Information Security from Stockholm University and an MSc in Security and Risk Management from the University of Leicester. He is also an ISACA Certified Information Security Manager (CISM).

His research interests are offensive cyberspace operations, deterrence and intelligence studies.

Bill Hutchinson was Foundation IBM Chair in Information Security, Director of SECAU (Security Research Centre), and Coordinator of the Information Operations and Security programmes in the School of Computer and Security Science at Edith Cowan University, Perth, Western Australia. From 2000 to 2010, he was the Chief Editor and founder of the *Journal of Information Warfare,* a member of the editorial board of the *Journal of the Australian Institute of Professional Intelligence Officers (AIPIO)*, and the Chair of the Western Australian chapter of the American Society for Industrial Security (ASIS). At present, Bill is investigating the concept of deceiving autonomous robots, writing a book on deception, serving as Guest Editor of the 20th anniversary edition of the *JIW*, and supporting researchers and course designers whenever he can.

Dr. Martti Lehto, (Military Sciences), Col (GS) (ret.) works as a Professor (Cyber Security) in the University of Jyväskylä. He has over 40 years' experience in C4ISR Systems in Finnish Defence Forces. Now, he is a cyber security and cyber defence researcher and teacher and the Pedagogical Director of the Cyber Security MSc. program. He is also adjunct professor in National Defence University in Air and Cyber Warfare. He has over 130 publications on the areas of C4ISR systems, cyber security and defence, information warfare, artificial intelligence, air power, and defence policy.

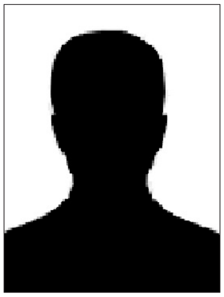
Christoph Lipps received a BSc. and MSc. in Electrical Engineering from the University of Kaiserslautern (TUK), Germany. Since 2015, he has been a Researcher at the German Research Center for Artificial Intelligence (DFKI), which is the biggest European AI research institution and is the birthplace of the "Industry 4.0" strategy. Meanwhile, he is a Lecturer and PhD candidate at the TUK as well. His research interests include Physical Layer Security (PhySec), Physically Unclonable Functions (PUFs); Artificial Intelligence (AI); identification and authentication of various entities, including biometric authentication of humans; as well as cyber security in general. In these areas, he is the author of about 20 scientific publications, has been in the TPC, has served as a reviewer for many conferences and journals, and has participated in a number of German and European research projects such as CoCos, IUNO and SCRATCh.

Sachinkumar Bavikatti Mallikarjun received a B.Eng from RV College of Engineering, India, in 2014 and an MSc. degree in Computer Science from the University of Kaiserslautern, Germany, in 2019, where he is currently pursuing a Ph.D. with the Institute for Wireless Communications and Navigation, Kaiserslautern. In 2019, he joined the Institute for Wireless Communications and Navigation as a Researcher. He has been involved in the 5G Modellregion Kaiserslautern project and has contributed to over 6 peer-reviewed publications. His research is in areas of Physical Layer Security, context awareness, mobility, and resource management of cellular networks.

Hans Dieter Schotten is Full Professor and Director of the Chair for Wireless Communication and Navigation of the University of Kaiserslautern. He is also a Scientific Director and member of the management board of the German Research Center for Artificial Intelligence (DFKI) where he heads the Department for Intelligent Networks. Before joining academia, he held industry positions in Ericsson and Qualcomm. Since 2018, he has been the Chairman of the German Information Technology Society ITG and a member of the supervisory board of the German VDE. His research interests are in mobile and industrial communications, network security, and AI. Hans Dieter Schotten received his Diploma and PhD in Electrical Engineering from the RWTH Aachen University in 1990 and 1997, respectively.

Jussi Simola has worked as a Cybersecurity specialist in Laurea University of Applied Sciences, and he is a PhD student in Cyber Security at University of Jyväskylä. His area of expertise includes decision support technologies, Situational Awareness Systems, information security, and continuity management. His current research focuses on development of the next generation Hybrid Emergency Response Model. He has also participated in the development of a common Early Warning System for the EU member countries.

Mathias Strufe received his diploma in Communications Engineering at the University of Applied Science (Kaiserslautern) in 2010 and in parallel a B.Eng. with First Class Honors in Electrical and Electronic Engineering from the University of East London. Since his start at DFKI in the Department of Intelligent Networks, he has been involved in several EU and nationally funded research projects with topics on AI based 5G network management in industrial environments. (mathias.strufe@dfki.de)



Douglas S. Wilbur, Ph.D. University of Missouri at Columbia, specializes in the role of propaganda and information warfare in violent conflict and war. In a previous life, he was an Information Operations Officer in the U.S. Army with four deployments.

Andrew Williams is an Australian Army Officer with extensive experience in strategy and capability, and a PhD student with the University of New South Wales Canberra, Cyber. He holds four master's degrees across science, arts, project management, and business administration. His main research areas are strategy, cybersecurity policy, and cyber law.

Richard L. Wilson is a Professor of Philosophy and Computer and Information Sciences at Towson University in Towson, MD, and Senior Research Scholar in the Hoffberger Center for Professional Ethics at the University of Baltimore, MD. Professor Wilson is a specialist in applied ethics with a variety of publications in cyber warfare ethics, information warfare ethics, and ethics of warfare. In addition, he works in business ethics, engineering ethics, environmental ethics, media ethics, and medical ethics. Teaching ethics in a wide variety of areas has led Professor Wilson to the commitment to an interdisciplinary approach to and in all fields of ethics and the centralizing in all of his work to "anticipatory ethics".

Evolution of Australia’s Cyber Warfare Strategy

AM Williams

University of New South Wales
Canberra, Australia

E-mail: andrew.williams3@student.adfa.edu.au

Abstract: *Since 2000, Australia has re-positioned itself from a country having scant recognition of cyber warfare to a nation with limited offensive and defensive capability facing increasing cyber incidents from at least one state-based actor (informally attributed as China). The dominance of a continental defence culture hindered the early development of a robust cyber warfare capability, resulting in a limited focus towards national infrastructure security. A transition to a far more robust, forward looking, strategic culture, dominated by statements of a perceived adversary in China, has enabled a greatly strengthened cyber warfare technical capability, though perhaps still lacking in personnel and supporting elements.*

Keywords: *Strategic Culture, Cyber Warfare, Information Warfare, Australia*

Introduction

The development of a robust and responsive Australian cyber-warfare capability has been delayed by Australia’s strategic culture, which has been dominated by a continental defence subculture since the 1970s. The use and potential impact of cyberspace as a warfighting domain has been evident since the late 1990s. The body of evidence over the past 20 years demonstrates the transition from the largely non-state threat of simple computer viruses to the incorporation of cyber warfare into military strategic campaigns by nations with more offensive strategic cultures than Australia, including the United States, Russia, and China.

The application of strategic culture theory is both a descriptive and predictive lens for policy analysis. Analysis of the historical record through a strategic culture lens enables a rich description of the military and political elites’ thinking regarding and contribution to capability development. Understanding the dominant strategic subculture of the day enables predictive analysis to predict where capability development may trend to in future iterations. Cyber warfare, like any military capability, is subject to the influence of strategic culture. This paper employs a fourth-generation definition of strategic culture and a broad definition of cyber warfare, incorporating cyber security of elements of national power through to offensive capabilities in order to determine the influence of strategic culture on Australia’s cyber warfare capability.

The paper considers publicly available Australian cyber warfare policy, capability realisation statements, and behaviours in the context of the global history of cyber warfare to determine whether a purely evidence-based approach to the formation of cyber policy has been adopted. Variation in policy and behaviour from the evidence base suggests Australia’s cyber thinking has been shaped by other factors, notably cultural ones. This paper is part of ongoing research into the influence of strategic culture on the development of Australia's cyber warfare capability since first publicly entering the field in 2000. The paper finds that Australia's dominant strategic subculture has contributed to a lag in early adoption of cyber warfare capability, and that the development of capability has remained reactive to Australia's dominant strategic subcultural influence.

Context

Strategic culture theory

Strategic culture theory lacks consistency and clarity, with ongoing debate surrounding the definition, constituent elements, and methodology, which poses opportunities and challenges to those seeking to employ it. This paper considers Australia's strategic subcultures and adopts Bloomfield's fourth generation definition: "the habits of ideas, attitudes, and norms toward strategic issues, and patterns of strategic behaviour, which are relatively stable over time" (Bloomfield 2011, p. 288).

Australia's strategic culture

Australia's strategic culture is predominantly comprised of competition between the subcultures of forward defence and continental defence (defensive neorealism), with elements of subordinate internationalism. Australia was founded on the concept of forward defence of the British Empire, which predominantly stood the test of time through two world wars. Australia's strategic culture rests on the concept of a 'great and powerful ally', originating with allegiance to the British Empire and transitioning to the U.S. between the end of WWII and the end of Australia's commitment to the Vietnam War. Following the Vietnam War, Australia transitioned to a period dominated by continental defence, theoretically optimised in the *1987 Defence White Paper* (Defence 1987). Prime Minister John Howard was elected in 1996, and oversaw a drift towards a forward defence culture (Bloomfield 2011), though the government's conservative position to reduce engagement in Asia, with a commensurate increase in the relationship with the United States, secured only a "minor shift in the relative prioritisation of the two primary sub-cultures, and continental defence remained dominant" (Bloomfield 2011, p. 234). The significant deployment to East Timor in 1999 dominated Australia's military thinking at the time, although this deployment is regarded by Bloomfield as 'a strategic *aberration* from the ordinary strategic cultural norm' (Bloomfield 2011, p. 106) rather than evidence of a return to forward defence.

Cyber warfare

There is a clear gap in the strategic culture literature regarding the study of cyber warfare. Gray (2013, p. 7) noted that "high-quality strategic theory about cyber simply is not there in the literature during the 1990s and most of the 2000s". As highlighted by Williams (2019), research on strategic culture and cyber warfare is lacking in consistency and depth and is hindered by a clear understanding of its constituent elements and an agreed methodology. There are a range of perspectives that may be considered in arriving at a definition of cyber warfare. Early consideration by Rid (2012) adopted a Clauswitzian argument to conclude that the lack of violence in cyber incidents prevents them from being considered as cyber 'war'. Discussing maritime warfare, which is akin to the global commons of cyberspace, Corbett (2004, p. 99) is clear that "anything... which we are able to achieve towards crippling our enemy's finance is a direct step to his overthrow". *The Tallinn Manual 2.0* takes the international law perspective that cyber psychological operations, interference with e-commerce, and funding hacktivist groups do not constitute a use of force (Schmitt 2017) although these actions may violate other rules of international law. Cyberspace actions in the grey zone, those that fall below the threshold of a use of force, are increasingly being employed by states to coerce and to influence. In the strictest legal sense, almost all actions in cyberspace do not rise above the threshold of a use of force, which accords with Rid's view that there will not (and cannot) be a cyber 'war'; however, the employment of cyber capability in the present day targets elements of national power which correspond with Corbett's theory of maritime warfare. Therefore, this article adopts a broad interpretation of cyber warfare incorporating Corbett's strategic theory such that cyber warfare includes any security, defensive, or offensive action in or through cyberspace with the intent of securing against the threat of cyber warfare or targeting a state or organisation's elements of (national) power. While such actions may not rise above the threshold of a use of force and thus may not be considered an act of conventional war, they may still be considered within the realm of cyber warfare.

Cyber warfare pre-2000

There is some conjecture about the advent and beginnings of cyber warfare. Alan Turing's WWII bomb was a computer-based effort to decipher modern encrypted communication, although it was more of a stand-alone computer to aid in a brute-force attack against the encryption device than what would now be considered cyber warfare. Brown (2017) argues that cyber warfare began with the advent of ARPANET and Ware's 1967 article on privacy and security of networks. The Trans-Siberian Pipeline explosion in 1982 is debated by Weiss (1996), Reed (2004), and Carr (2012) as a cyber warfare event, with no agreement reached. Further, Kaplan (2016) provides detail on US J39 operations in Serbia in the late 1990s, including cyberattacks against Serbian telephone and air defence networks.

Moonlight Maze is the first widely reported and acknowledged cyber incident. This was a sophisticated Russian intelligence operation aimed at unclassified yet sensitive U.S. military and other sites, detected in early 1998 (Adams 2001), and made public by the FBI in 1999 (Elkus 2013). The evidence base prior to 2000 was such that Australia and other nations were in a position to be informed and act.

Development of Australia's Cyber Warfare Capability

2000-2008: Defensive rhetoric

Australia's first tentative step into cyber warfare policy statements occurred in 2000. The *2000 Defence White Paper* (Defence 2000) stated that Australia faced nonmilitary threats, including cyberattack, and that Defence would contribute to the solution. There is no publicly available evidence to suggest that resources were made available to enact this policy statement with the 2001 Annual Report merely acknowledging the statement in the white paper, although strengthening the language to "cyber warfare" (Defence 2001, p. 4). In 2002, *The Australian approach to warfare* (Defence 2002a), and *Force 2020* (Defence 2002b), make no reference to cyber warfare. Similarly, the two subsequent updates to the Defence white paper (Defence 2003, 2005) fail to mention cyber.

Concurrent with Australia's policy apathy towards cyber warfare, China was actively targeting U.S. and British computer networks, under an operation dubbed Titan Rain by the FBI. Publicly revealed in 2005 ('Titan Rain' 2005), Kaplan (2016), Segal (2013) and Whitmore (2016) outline the severity and widespread nature of these incidents, potentially first seen at Lockheed Martin in 2003. Segal (2013) further alleges Chinese espionage activities continued under what was dubbed Operation Byzantine Hades. This increasing body of evidence of persistent threats in cyberspace should have influenced policy and action in Australian approaches to cyber warfare. The *2000 Defence White Paper*, subsequent updates, and aforementioned policy statements maintained the dominant defensive strategic culture position, which minimised the threat. This approach ignored the growing evidence base and precluded development of an Australian cyber warfare capability beyond basic cyber security.

The 2007 update to the *2000 White Paper* acknowledges "an emerging need to focus on 'cyber-warfare', particularly capabilities to protect national networks to deny information" (Defence 2007a, p. 53). This update provides a shift from the focus on nonmilitary threats in 2000 to the concept of cyber warfare, although it remained focussed on national infrastructure, perhaps in response to Titan Rain and Byzantine Hades. *Joint Operations for the 21st Century* (Defence 2007b), released two months earlier, neglected to mention cyber at all, suggesting that remarks in the 2007 update were focussed entirely on national security. Again, this perpetuated the continental defence strategic culture and a lack of comprehension of the military application of cyber warfare.

While Australia made limited statements regarding cyber security of national infrastructure, the nature of cyber warfare continued to evolve. In 2007, alleged Russian backed hackers *attacked* Estonian Internet-based services systems. This incident is particularly important as it became evident to the public that cyberspace could be used in international conflict (Schmidt, A 2013). Also, in 2007, Operation Orchard marked a transition to the use of cyber warfare in concert with physical military actions (although, as mentioned, similar employment may have occurred approximately a decade earlier via US J39 operations in Serbia). This combined action between the U.S. and Israel saw the Israeli Air Force penetrate Syrian airspace after cyberattacks degraded Russian-produced air defence systems.

Significant cyber incidents continued in 2008. A crude oil pipeline explosion in Turkey preceded the Russian invasion of Georgia. The incident was attributed to Russian-backed hackers (Whitmore 2016) and was potentially the first physical destruction arising from a cyber weapon. Similar to Operation Orchard, the Russian invasion of Georgia in 2008 was accompanied by large-scale cyber incidents, highlighting the ability to coordinate cyber effects within a military campaign. Further, the vulnerabilities revealed through detection of a malicious payload on U.S. military systems in the Middle East was said by Grindal (2013) and Whitmore (2016) to have led to the creation of the US Cyber Command in 2010. Meanwhile, despite Australia being a part of coalition operations in the Middle East, Australia's policy position on cyber warfare up to 2008 remained underdeveloped and inadequate to meet this changing military landscape.

The *National Security Speech* made by the newly elected Prime Minister Rudd (House of Representatives 2008) indicated an upcoming white paper (released in 2009) would address "the emergence of new challenges … [including] new threats such as cyber warfare" (Commonwealth of Australia 2008, p. 12552). The speech subsequently refers to the increasing reliance on information networks and vulnerability to "cyber attacks that may disrupt the information that increasingly lubricates our economy and system of government" (Commonwealth of Australia 2008, p. 12553). Despite a decade of escalating incidents in cyberspace, the government continued to refer to the "emergence" of the threat, contrary to the increasing evidence of cyber events occurring across the globe. While Defence sought a more balanced approach to enable strategic options around this period (Bloomfield 2011), the dominance of continental defence culture had not sufficiently declined to enable proper recognition of the need for new military capabilities, such as cyber warfare.

The policy statements over the period 2000-2008 are marked by indifference to cyber warfare and references to the protection of national infrastructure rather than the potential military applications in cyberspace. Australian strategic culture in relation to cyber warfare thus remained not only defensive and reactive but was also increasingly lagging the reality of world events. It is difficult to assume that any relevant actions were taken within the department during this period as Defence annual reports largely do not refer to cyber. Further, the policy statements regarding cyber were (as noted) limited to national infrastructure cyber security. Despite rapid increases in cyber warfare incidents, the Australian government maintained a reactive, defensive posture during this period.

2009–2015: Recognition of the need, alliances, and classified development

The transition to the second period in the development of Australia's cyber warfare capability was marked by the provision of a national strategy and an increasing number of relevant statements in the Defence portfolio. The national *Cyber Security Strategy* and a new Defence white paper, *Defending Australia in the Asia Pacific Century: Force 2030,* were released in 2009. The *Strategy* (Attorney General 2009) lacks any reference to cyber warfare, maintaining the national perspective that cyber poses a threat only in terms of infrastructure security. The discussion of Defence's role is limited to the establishment of the Cyber Security Operations Centre (CSOC) within the Defence Signals Directorate, and the Directorate's role more broadly (Attorney General 2009). The *2009 White Paper* maintained a continental defence position, retreating from Howard's drift to forward defence. It makes bold statements regarding the rise of cyber warfare, placing it within the top band of capabilities required alongside undersea and anti-submarine warfare, air superiority, and strategic strike capabilities (Defence 2009b). The overture from the Defence Minister reiterates previous rhetoric of the threat to critical infrastructure and continues the narrative of cyber warfare as an "emerging area" (Defence 2009b, p. 14).

While Australia continued to ponder the 'emergence' of cyber warfare, its primary ally, the U.S., was allegedly building the most important weapon in the history of cyber warfare to date. The STUXNET attack against the Iranian nuclear processing facility in Natanz provided the first confirmed evidence of a cyberattack producing physical effects. Given the sensitivity of the development of such a weapon, there is potential that the Australian government would not have been aware of the attack prior to public awareness in 2010, when it was discovered by the infosec community (Fruhlinger 2017). Its emergence should have generated considerable debate, policy, and outcomes. The Australian policy and strategy community, however, remained docile.

Defence Annual Reports following the *2009 White Paper* provide little by way of demonstrable outcomes. 'Cyber' is mentioned in the 2009 Defence Annual Report (Defence 2009a) (the first mention in an annual report since 2000/01), the first real reflection that policy was having any impact or resulting in any recognition of the need for action. The Department of the Prime Minister and Cabinet Report 2010-11 provides little change other than to note the transfer of cyber policy to the Department of the Prime Minister and Cabinet (PM&C). In turn, PM&C flagged "development of the first Cyber White Paper and a risk framework to inform national security community decision making" (Prime Minister and Cabinet 2011, p. 8). The Cyber White Paper was not delivered until the release of the 2016 Cyber Security Strategy, perhaps more as a result of the multiple changes in Prime Ministers (Gillard, Rudd, Abbott, Turnbull) than reflecting ongoing apathy to the development of new policy.

A significant policy event occurred in late 2011 to strengthen an alliance approach to cyber warfare. Colloquially known as ANZUS 2.0, the Joint Communique released at the conclusion of the *Australia-United States Ministerial Consultations* (AUSMIN) stated that "in the event of a cyber attack…Australia and the United States would consult together and determine appropriate options to address the threat" (Foreign Affairs and Trade 2011). The statement was criticised by some, believing that it raised more questions than it resolved, and that the threshold for consultation and action was too high (Davies *et al.* 2012). This development does, however, speak to Australia's strategic policy intent of maintaining relationships with great and powerful allies. It also perhaps marks a step towards a forward defence posture, although with continental defence retaining primacy. Australian Departmental reporting provided a series of statements regarding cyber (Defence 2012; Foreign Affairs and Trade 2012; Prime Minister and Cabinet 2012), criticised by Thompson (2012, p. 65) as demonstrating "a disjointed and, at times, confused approach to policy", while also calling for an offensive cyber capability.

In 2013, Prime Minister Gillard announced the creation of a new Australian Cyber Security Centre (ACSC) as the "hub of the government's cyber security efforts" (Gillard 2013). While not offering new capability, this transition sought to bring together disparate elements for a more cohesive approach. A new Defence *White Paper* (2013) also changed the name of the Defence Signals Directorate to the Australian Signals Directorate (ASD). The same paper signalled an expanded alliance relationship, incorporating Australia, the U.S., and the UK, and indicated, for the first time, that cyber capabilities would be integrated "into routine planning and command and control processes" (Defence 2013, pp. 78-9). This statement indicates that cyber capabilities were being incorporated into military operations, perhaps marking the transition to a broader cyber warfare capability. Feakin (2013) takes a different view, believing that the loss of the term 'warfare' associated with cyber indicated "an attempt to de-militarise" the issue. While the policy statements increasingly referred to cyber, the output statements suggest limited translation of that rhetoric into capability outcomes. It appears that the dominant continental defence posture may have continued to hinder capability development.

In 2015, ASD detected an intrusion into the Australian Bureau of Meteorology (BoM). The Government report mentions the breach "had also been used to compromise other Australian government networks" and attributed the incident to "state-sponsored cyber adversaries" (ACSC 2016, p. 11). The report also noted that the Department of Parliamentary Services had been compromised, although it does not elaborate further. This was the first public acknowledgement that the Australian Government had been the victim of state-sponsored cyber incidents. While not rising to the level of a use of force, these incidents indicated that Australia was not immune.

Australia appeared to retain a defensive, reactive posture over this period, despite considerable evidence that other actors were readying themselves for cyber warfare on a grand scale. Lehmann (2015) argued for an offensive cyber capability, evoking many strategic cultural arguments for a return to a forward defence posture. Indeed, the title of his paper (incorporating the phrase "beyond the Maginot mentality") is a specific call to move beyond a defensive posture. Further, he highlights the cultural divide with statements against Asian 'others', while reinforcing the culturally-aligned U.S. cyber strategy. His paper notes previous policy initiatives—at least as announced in public—as being almost purely defensive in nature, while concurrently citing a small number of papers published between 2013 and 2015 advocating for offensive capabilities. This hinted that the tide was turning, if it had not already.

While national statements were publicly focussed on defensive security, something was nonetheless changing, hidden behind the veil of national security. As explained in the next section, an offensive cyber warfare capability was at least in development, if not already being employed.

2016: Acceleration, declaration, and operations

This period in Australian cyber warfare capability is marked by a series of significant announcements in 2016. The government declared that it possessed an offensive cyber capability (Turnbull 2016) and that it had been deployed on operations (House of Representatives 2016). It is apparent that Australia must have possessed an offensive capability prior to this epoch, and that it was developed and implemented in secrecy. It is unlikely to have occurred prior to the *2009 White Paper* and was perhaps linked to the increased alliance framework for cyber operations between 2011 and 2013, with employment possibly aligned with the operations-planning processes and approvals noted in the *2013 White Paper*. Concurrent with the announcement of the operational employment of an offensive capability, the government adopted 'cyber storm' theory (Tehan 2016), flagging the potential for considerable concurrent malicious actions in cyberspace. Despite such declarations, Thompson (2016, p. 47) indicates the state of Australian Defence Force (ADF) cyber capability was inadequate, arguing for "an appropriate policy and legal framework" and the allocation of resources "for the education, training, and equipping of cyber warriors". His analysis suggests ADF operational capability was not appropriately resourced beyond technical systems for security, defensive, and offensive operations.

The February *2016 Defence White Paper* was silent on the question of offensive cyber capability. It does, however, refer to an ability to "deter" (Defence 2016a, p. 18), with deterrence a key point within the Prime Minister's announcement of an offensive capability. The decision to retain secrecy regarding the offensive capability in a major policy statement, while publicly announcing it shortly, thereafter, suggests a degree of uncertainty in the elite. Such a statement would be aligned with a more forward defence posture, while this *White Paper* refused to prioritise any position with equally weighted Strategic Defence Objectives. The associated *Integrated Investment Program* provides a "cyber security capability improvement" funding line of $300-400m between 2016 and 2025 (Defence 2016b), and a workforce of 900 military and 800 Australian Public Service positions across intelligence, surveillance, reconnaissance, electronic warfare, space, and cyber domains over the decade (Defence 2016b). Despite the significant announcements and the increasing reference in the *White Paper*, there is no specific reference to the development of the offensive cyber capability in the *Defence Annual Report 15-16*, nor significant change in reporting (Defence 2016c). The uncertainty is in the elite perhaps playing out between the political and military elements.

In 2017, Prime Minister Turnbull announced he had authorised ASD to use offensive capabilities against offshore criminals (Turnbull 2017). The 2017 *International Cyber Engagement Strategy* (Foreign Affairs and Trade 2017) provided the clear statement of Australia's division of labour between the Department of Defence and the ADF; that the offensive capability was resident in ASD only; and that ADF operations were executed by ASD under direction of the Chief of Joint Operations, relying on the legislative powers of ASD rather than executive powers under a royal war prerogative (Evans & Williams 2019). This division of labour retains high-end capability in a central repository, able to act against terrorists, criminals, and military targets alike. What remains unclear from this position is the depth of the capability and which 'client' ASD will be prioritised to support in the event of multiple competing tasks, for example, a cyber storm.

The establishment of the Information Warfare Division in July 2017 is the key example of readiness to adopt cyber warfare as a conventional capability alongside more traditional naval and military capabilities, and a clear indicator of how long it has taken the ADF to take cyber warfare seriously. This was closely followed by the creation of the Defence SIGINT and cyber command in January 2018 (Defence 2018), preceding the establishment of ASD as a statutory authority on 1 July 2018. The Australian Strategic Policy Institute (ASPI) remained sceptical of the resourcing and Defence's ability to build the required capability (Paterson 2018). The Head of the Information Warfare Division, Major General Thompson, shared concern about the nation's ability to act in the event of a significant cyber incident (Borys 2019), supporting ASPI's scepticism. Despite the considerable number of announcements in 2016, and the allocation of significant resources over the decade 2016-2025 (including the raising of an Information Warfare Division and the SIGINT and cyber command) senior leadership remained concerned about the ability to defeat a major cyberattack, such as a cyber storm, against Australia. Air Marshal McDonald stated to the Australian Senate that as of late 2019, "Defence is well on the way to developing a defensive cyber capability", that a deployable defensive cyber project would be considered by government in the near term, and that by 2023 there will be 446 personnel for cyber defence (Senate 2019, pp. 53-4, 76). This reporting appears to be exclusive of personnel and projects within ASD, reflecting their independence as a statutory authority.

The rise of the Information Warfare Division and SIGINT and cyber command occurred at a time when there were both increasing rhetoric against an expansive China and cyber incidents against Australia. Claims of Chinese influence on the political landscape and of espionage are increasing; with Jennings (2017) citing Senator Dastyari's fall from grace along with funding to business, universities, and political donations as prime examples of China's reach into Australia. Further, two significant breaches of the Australian National University systems, in 2018 and 2019, have been loosely attributed to China (Crozier 2019; Schmidt, B 2019). In early 2019, Australia had its first "national cyber crisis", a "malicious intrusion into the Australian Parliament House computer network" (Australian Signals Directorate 2019). While not formally attributed (Senate 2019), China has been unofficially held responsible (Vincent 2019). The expansion of rhetoric towards China may be attributed to more forward-leaning policy and behaviour in Australia's cyber warfare capabilities. Ultimately, the increasing cultural rhetoric has enabled the declaration of an offensive cyber capability. The development of the military defensive capability continues to evolve following the 2016 resource allocations.

The most recent announcements by the Australian Government, based on strategic analyses conducted primarily in 2019, include cyber warfare packages specific to the ASD and Defence more broadly. The Prime Minister's media announcement of $1.35 billion and 500 positions for ASD over the coming decade (Morrison 2020a), to be documented in the forthcoming updated national cyber strategy, was announced the day prior to the *2020 Defence Strategic Update* (Defence 2020a) and associated 2020 Force Structure Plan (Defence 2020b). Prior to the policy announcements, a combined media release from the PM, the Minister for Home Affairs, and the Minister for Defence announced that Australia was under cyberattack from a malicious state-based actor, although the release did not specifically attribute the state involved (Morrison 2020b). Media were quick to cite "senior sources" as saying that the perpetrator was China (Hitch & Probyn 2020). The media release in mid-June served to forewarn and to establish credence for the cyber announcements and broader defence funding. The update reflects increased rhetoric against China, stating that there is competition primarily between the U.S. and China, being played out in the Indo-Pacific and in Australia's immediate region. This language serves to deflect competition away from the Australia/China relationship. Notwithstanding, the update is more aggressive in tone, pointing to competition, coercion, malicious activities in the grey zone, and the threat to rules, norms and institutions. The update is geographically narrowing Australia's strategic focus to the Indo-Pacific, though it is not a return to a dominant continental defence strategic subculture. The fundamental difference is the ever-increasing volume and level of negative statements regarding China. This policy position contrasts with Bloomfield's (2011) theoretical construct of continental defence which requires a lack of stated adversary, and demonstrates alignment with Davis' (2018a, 2018b, 2018c, 2019) repeated calls for "forward defence in depth", which is also the effective conclusion in Dibb's (2020) analysis. The funding announcement for "information and cyber capabilities" in the update is for $15 billion over the decade, which is a significant increase on previous funding, though only represents 6% of the investment plan, which pales into insignificance beside the traditional physical domains which each receive between 20-28% of the funding. The update acknowledges the increasing competition in the grey zone yet maintains development of traditional tangible capabilities that are less capable of strong effect in the grey zone. The release of the updated national cyber strategy (Home Affairs 2020) in August 2020 made little change beyond the announcements of the previous two months.

Limitations and Future Direction

This article is part of a body of ongoing PhD research being conducted at the University of New South Wales Canberra Cyber, following the methodology described in Williams (2019). It articulates the influence of Australia's strategic culture on development of its cyber warfare capability. Analysis of artefacts and behaviour in the context of the prevailing strategic culture provides opportunities for an understanding of bias for policy makers, particularly when comprehending Australia's policy and behaviour in comparison to others.

Further work is required to develop a richer understanding of the development of the cyber warfare capability. Further analysis of primary sources in the form of public statements made by the elite and by executives may provide additional information regarding cultural views and development of cyber warfare capability. The time at which Australia began to develop its offensive cyber warfare capability requires further analysis, particularly when considering the events that may have led to a change in policy. This field is evolving and will continue to do so over the coming years as Australia's cyber warfare capability is embedded as a part of the total force rather than being viewed as a niche capability.

Conclusion

The emergence of cyberspace from being an academic research tool to a dual use technology with increasing military application is unsurprising. Australia's inability to maintain pace with the emergence of cyber warfare as a threat and to develop its own cyber warfare capability is attributable to the prevailing strategic subculture of continental defence. The focus of capability development has predominantly vested in security for elements of national power, enabling a degree of protection against adversary operations in the grey zone. This has come at the expense of contributions to traditional offensive and defensive warfare capability. Some 20 years after the first appearance of cyber in a Defence white paper, the project for a deployable military defensive cyber capability is about to be presented to government for approval and, by 2023, fewer than 500 ADF personnel will be available for cyber defence. Australia has, however, developed an offensive capability which is resident in the nationally focussed ASD. Notwithstanding, that offensive capability may be employed in military operations by the SIGINT and cyber command operating under the legislative umbrella of ASD.

Despite early publicly available evidence of the nature of the threat, Australia's opening foray into cyber policy was a single statement resulting in seeming inaction. Such scant recognition of the increasing challenge lasted almost a decade, until the release of a national strategy in 2009. As late as the establishment of U.S. Cyber Command in 2010, Australia was still referring to cyber warfare as an emerging threat. The period from 2009 to 2015 saw a transition from inaction to an understanding of the threat and development of Australia's nascent defensive and offensive cyber warfare capability. Realisation of cyber warfare capability was slow and was often accompanied by the shuffling of responsibilities between government departments rather than true advances.

Beyond the national security element, Australia's fledgling cyber warfare capability has advanced in the period from 2016. The declaration of ASD's offensive capability and the establishment of ADF organisations marks a transition to realising the policy statements made almost two decades prior, and enabling a deterrence posture. Capability realisation is occurring concurrently with increases in strategic cultural statements of alignment with traditional Western allies, and growing statements against those presumed to be strategic cultural 'others', namely China, who have been informally connected to cyberattacks against Australia.

Ongoing work in this field will contribute additional insights into the development of Australia's cyber warfare capability. A deeper understanding of Australia's strategic culture enables identification of biases and weaknesses in policy and capability realisation, enabling a more robust capability. Cyber warfare capability has lagged behind the global evidence base because the dominant strategic subculture of continental defence has denied acknowledgement of adversaries in cyberspace. The transition to forward defence has enabled growth in cyber warfare capability, although more is needed to meet the growing issue of state-based actor interference in Australia's sovereignty.

Note

This paper does not represent the views of the Australian Government nor the Australian Department of Defence.

References

Adams, J 2001, 'Virtual defense', *Foreign Affairs*, vol. 80, no. 3, pp. 98-112.

Attorney General, Department of 2009, *Cyber Security Strategy*, Commonwealth of Australia, viewed 13 February 2020, <https://www.ag.gov.au/RightsAndProtections/CyberSecurity/Documents/AG%20Cyber%20Security%20Strategy%20-%20for%20website.pdf>.

Australian Cyber Security Centre (ACSC) 2016, *2016 Threat Report*, Commonwealth of Australia, viewed 13 February 2020, <https://www.acsc.gov.au/publications/ACSC_Threat_Report_2016.pdf>.

Australian Signals Directorate 2019, *Australian Signals Directorate Annual Report 2018-19*, Commonwealth of Australia, viewed 13 February 2020, <https://www.asd.gov.au/sites/default/files/2019-10/annual_report_2018-19.pdf>.

Bloomfield, A 2011, 'Australia's strategic culture: An investigation of the concept of strategic culture and its application to the Australian case', Doctor of Philosophy thesis, Queen's University.

Borys, S 2019, *Senior Defence figure raises concerns about future cyber attacks - and the scenario costing him sleep*, ABC News, viewed 16 November 2019, <https://www.abc.net.au/news/2019-02-19/australian-army-under-cyber-attack-major-general-marcus-thompson/10822966>.

Brown, J 2017, 'Cyberwar isn't so new—It began in 1967', *Army*, vol. 67, no. 9, pp. 65-7.

Carr, J 2012, 'The myth of the CIA and the Trans-Siberian Pipeline Explosion', viewed 13 August 2019, <http://jeffreycarr.blogspot.com/2012/06/myth-of-cia-and-trans-siberian-pipeline.html>.

Corbett, JS 2004, *Principles of maritime strategy*, Dover Publications, New York, NY, US.

Crozier, R 2019, 'ANU suffers second "significant" hack in a year', viewed 17 November 2019, <https://www.itnews.com.au/news/anu-suffers-second-significant-hack-in-a-year-526123>.

Davies, A, Lewis, J, Herrera-Flanigan, J & Mulvenon, J 2012, *ANZUS 2.0: Cybersecurity and Australia - US relations*, no. 46, ASPI.

Davis, M 2018a, '"Forward defence in depth" for Australia (part 1)', *The Strategist*, viewed 14 July 2020, <https://www.aspistrategist.org.au/forward-defence-in-depth-for-australia-part-1/>.

——2018b, '"Forward defence in depth" for Australia (part 2)', *The Strategist*, viewed 14 July 2020, <https://www.aspistrategist.org.au/forward-defence-in-depth-for-australia-part-2/>.

——2018c, ''Forward defence in depth' for Australia (part 3)', *The Strategist,* viewed 14 July 2020, <https://www.aspistrategist.org.au/forward-defence-in-depth-for-australia-part-3/>.

——2019, '"Forward defence in depth" for Australia', *Strategic Insights*, vol. 139, viewed 14 July 2020, <https://s3-ap-southeast-2.amazonaws.com/ad-aspi/2019-06/SI%20139%20Forward%20defence%20in%20depth.pdf?vfiVknPEa5saKIFEqC_jI_5IFSkwtKvg>.

Defence 1987, *The Defence of Australia*, Commonwealth of Australia, viewed 15 July 2020, <https://www.defence.gov.au/Publications/wpaper1987.pdf>.

——2000, *Defence 2000: Our future Defence Force*, Commonwealth of Australia, viewed 13 February 2020, <http://www.defence.gov.au/publications/wpaper2000.pdfhttp://www.defence.gov.au/publications/wpaper2000.pdf>.

——2001, *Defence Annual Report 2000-01*, Commonwealth of Australia, viewed 13 May 2020, <https://www.defence.gov.au/AnnualReports/00-01/full.pdf>.

——2002a, *The Australian approach to warfare*, Commonwealth of Australia, viewed 13 February 2020, <https://defence.gov.au/publications/docs/taatw.pdf>.

——2002b, *Force 2020*, Commonwealth of Australia, viewed 13 February 2020, <https://www.defence.gov.au/publications/f2020.pdf>.

——2003, *Australia's national security: A defence update 2003*, Commonwealth of Australia, viewed 13 February 2020, <http://aseanregionalforum.asean.org/files/library/ARF%20Defense%20White%20Papers/Australia-2003.pdf>.

——2005, *Australia's national security: A defence update 2005*, Commonwealth of Australia, viewed 13 February 2020, <http://www.dtic.mil/dtic/tr/fulltext/u2/a481121.pdf>.

——2007a, *Australia's National Security: A Defence Update 2007*, Commonwealth of Australia, viewed 13 February 2020, <http://www.operationspaix.net/DATA/DOCUMENT/944~v~Australia__National_Security.pdf>.

——2007b, *Joint Operations for the 21ST Century*, Commonwealth of Australia, viewed 13 February 2020, <https://www.defence.gov.au/publications/fwc.pdf>.

——2009a, *Defence Annual Report 2008-09*, vol. 1, Commonwealth of Austraia, viewed 13 February 2020, <https://www.defence.gov.au/AnnualReports/08-09/2008-2009_Defence_DAR_v1full.pdf>.

——2009b, *Defending Australia in the Asia Pacific Century: Force 2030*, Commonwealth of Australia, viewed 13 February 2020, <http://www.defence.gov.au/whitepaper/2009/docs/defence_white_paper_2009.pdf>.

——2012, *Department of Defence Annual Report 2011-2012*, Commonwealth of Australia, viewed 13 February 2020, <https://defence.gov.au/AnnualReports/11%2D12/>.

——2013, *White Paper 2013*, Commonwealth of Australia, viewed 13 February 2020, <http://www.defence.gov.au/whitepaper/2013/docs/WP_2013_web.pdf>.

——2016a, *2016 Defence White Paper*, Commonwealth of Australia, viewed 13 February 2020, <http://www.defence.gov.au/whitepaper/Docs/2016-Defence-White-Paper.pdf>.

——2016b, *2016 Integrated Investment Program*, Commonwealth of Australia, viewed 13 February 2020, <http://www.defence.gov.au/whitepaper/Docs/2016-Defence-Integrated-Investment-Program.pdf>.

——2016c, *Defence Annual Report 2015-16*, vol. 1, Commonwealth of Australia, viewed 13 February 2020, <https://defence.gov.au/AnnualReports/15-16/Downloads/DAR_2015-16_Vol1.pdf>.

——2018, 'Defence Chief announces new command', Commonwealth of Australia, viewed 13 February 2020, <https://news.defence.gov.au/media/media-releases/defence-chief-announces-new-command>.

——2020a, *2020 Defence Strategic Update*, Commonwealth of Australia, viewed 2 July 2020, <https://www.defence.gov.au/StrategicUpdate2020/docs/2020_Defence_Strategic_Update. pdf>.

——2020b, *2020 Force Structure Plan*, Commonwealth of Australia, viewed 2 July 2020, <https://www.defence.gov.au/StrategicUpdate-2020/docs/2020_Force_Structure_Plan.pdf>.

Dibb, P 2020, 'Is Morrison's strategic update the defence of Australia doctrine reborn?', *The Strategist*, viewed 14 July 2020, <https://www.aspistrategist.org.au/is-morrisons-strategic-update-the-defence-of-australia-doctrine-reborn/>.

Elkus, A 2013, 'Moonlight maze', *A fierce domain: Conflict in cyberspace, 1986 to 2012*, ed. J Healey, CCSA, Vienna VA, US, pp. 152-63.

Evans, H & Williams, A 2019, 'ADF offensive cyberspace operations and Australian domestic law: Proprietary and constitutional implications', *Federal Law Review*, vol. 47, no. 4, pp. 606-30.

Feakin, T 2013, 'What's in a name change? Cyber in the Defence White Paper', *The Strategist*, viewed 19 November 2017, <https://www.aspistrategist.org.au/whats-in-a-name-change-cyber-in-the-defence-white-paper/>.

Foreign Affairs and Trade, Department of 2011, *Australia-United States Ministerial Consultations 2011 Joint Communique*, Commonwealth of Australia, viewed 13 February 2020, <https://dfat.gov.au/geo/united-states-of-america/ausmin/Pages/ausmin-joint-communique-2011.aspx>.

——2012, *Department of Foreign Affairs and Trade Annual Report 2011-12*, Commonwealth of Australia, viewed 13 February 2020, <https://nla.gov.au/nla.obj-801640231/view?partIdnla.obj-802014908#>.

——2017, *Australia's International Cyber Engagement Strategy*, Commonwealth of Australia, viewed 13 February 2020, <http://dfat.gov.au/international-relations/themes/cyber-affairs/aices/pdf/DFAT%20AICES_AccPDF.pdf>.

Fruhlinger, J 2017, 'What is Stuxnet, who created it and how does it work?', *CSOonline*, viewed 13 February 2020, <https://www.csoonline.com/article/3218104/what-is-stuxnet-who-created-it-and-how-does-it-work.html>.

Gillard, J 2013, *Australian Cyber Security Centre*, Commonwealth of Australia, viewed 13 February 2020, <https://pmtranscripts.pmc.gov.au/release/transcript-19008>.

Gray, CS 2013, *Making strategic sense of cyber power: Why the sky is not falling*, Strategic Studies Institute, US Army War College Press, viewed 13 February 2020, <https://ssi.armywarcollege.edu/pdffiles/ PUB1147.pdf>.

Grindal, K 2013, 'Operation BUCKSHOT YANKEE', *A fierce domain: Conflict in cyberspace, 1986 to 2012*, ed. J Healey, CCSA, Vienna VA, US, pp. 205-11.

Hitch, G & Probyn, A 2020, 'China believed to be behind major cyber attack on Australian governments and businesses', ABC News, viewed 14 July 2020, <https://www.abc.net.au/news/2020-06-19/foreign-cyber-hack-targets-australian-government-and-business/12372470>.

Home Affairs, Department of 2020, *Australia's Cyber Security Strategy 2020*, Commonwealth of Australia, viewed 6 August 2020, <https://www.homeaffairs.gov.au/cyber-security-subsite/files/cyber-security-strategy-2020.pdf>.

House of Representatives 2008, *Parliamentary debates: House of Representatives National Security Speech*, Commonwealth of Australia, viewed 13 February 2020, <http://parlinfo.aph.gov.au/parlInfo/genpdf/chamber/hansardr/2008-12-04/0045/hansard_frag.pdf;fileType=application%2Fpdf>.

——2016, *Parliamentary Debates*, vol. 6, Commonwealth of Australia, viewed 13 February 2020, <http://parlinfo.aph.gov.au/parlInfo/download/chamber/hansardr/9b169b3b-768b-49e5-aabb-ed05f3ce0ebb/toc_pdf/House%20of%20Representatives_2016_11_23_4599_Official. pdf;fileType=application%2Fpdf#search=%22chamber/hansardr/9b169b3b-768b-49e5-aabb-ed05f3ce0ebb/0000%22>.

Jennings, P 2017, ''At last, we're awake to China's predatory meddling', viewed 13 February 2020, <https://www.aspi.org.au/opinion/last-were-awake-chinas-predatory-meddling>.

Kaplan, F 2016, *Dark Territory: The secret history of cyber war*, Simon & Schuster, New York, NY, US.

Lehmann, M 2015, 'The case for an offensive ADF cyber capability: Beyond the Maginot mentality', *Australian Defence Force Journal*, no. 198, pp. 31-8.

Morrison, S 2020a, *Nation's largest ever investment in cyber security*, Prime Minister of Australia, viewed 14 July 2020, <https://www.pm.gov.au/media/nations-largest-ever-investment-cyber-security>.

Morrison, S 2020b, *Statement on malicious cyber activity against Australian networks*, Prime Minister of Australia, viewed 14 July 2020, <https://www.pm.gov.au/media/statement-malicious-cyber-activity-against-australian-networks>.

Paterson, T 2018, 'The ADF's Information Warfare Division needs more staff and a clear framework', *The Strategist*, viewed 16 November 2019, <https://www.aspistrategist.org.au/the-adfs-information-warfare-division-needs-more-staff-and-a-clear-framework/>.

Prime Minister and Cabinet, Department of 2011, *Department of the Prime Minister and Cabinet Annual Report 2010-11*, Commonwealth of Australia, viewed 13 February 2020, <https://www.pmc.gov.au/sites/default/files/publications/annual_report_10_11.pdf>.

——2012, *Department of the Prime Minister and Cabinet Annual Report 2011-12*, Commonwealth of Australia, viewed 13 February 2020, <https://pmc.gov.au/sites/default/files/ publications/annual_report_11_12.pdf>.

Reed, TC 2004, *At the abyss: An insider's history of the Cold War*, 1st edn., Ballantine, New York, NY, US.

Rid, T 2012, 'Cyber war will not take place', *Journal of Strategic Studies*, vol. 35, no. 1, pp. 5-32.

Schmidt, A 2013, 'The Estonian Cyberattacks', *A fierce domain: Conflict in cyberspace, 1986 to 2012*, ed. J Healey, CCSA, Vienna, VA, US, pp. 174-93.

Schmidt, B 2019, 'Message from the Vice-Chancellor', viewed 17 November 2019, <https://www.anu.edu.au/news/all-news/message-from-the-vice-chancellor#overlay-context=user>.

Schmitt, MN (ed.) 2017, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, 2 edn., Cambridge University Press, Cambridge, UK.

Segal, A 2013, 'From TITAN RAIN to BYZANTINE HADES: Chinese cyber espionage', *A fierce domain: Conflict in cyberspace, 1986 to 2012*, ed. J Healey, CCSA, Vienna VA,US, pp. 165-73.

Senate 2019, *Foreign Affairs, Defence and Trade Legislation Committee Estimates*, Commonwealth of Australia, viewed 13 February 2020, <https://parlinfo.aph.gov.au/parlInfo/download/committees/estimate/53068544-efe7-4494-a0f2-2dbca4d2607b/toc_pdf/Foreign%20 Affairs,%20Defence%20and%20Trade%20 Legislation%20Committee_2019_10_23_7285.pdf;-fileType=application%2Fpdf#search=%22committees/estimate/53068544-efe7-4494-a0f2-2dbca4d2607b/0000%22>.

Tehan, D 2016, *Address to the National Press Club – 'A Cyber Storm'*, Department of the Prime Minister and Cabinet, viewed 1 October 2018, <https://ministers.pmc.gov.au/tehan/2016/address-national-press-club-cyber-storm>.

Thompson, M 2012, 'The cyber threat to Australia', *The Australian Defence Force Journal*, no. 188, pp. 57-70.

——2016, 'The ADF and cyber warfare', *The Australian Defence Force Journal*, no. 200, pp. 43-8.

'Titan Rain' 2005, viewed 12 January 2020, <https://www.cfr.org/interactive/cyber-operations/titan-rain>.

Turnbull, M 2016, *Launch of Australia's cyber security strategy Sydney*, Prime Minister of Australia, viewed 4 December 2016, <https://www.malcolmturnbull.com.au/media/launch-of-australias-cyber-security-strategy>.

——2017, *Offensive cyber capability to fight cyber criminals*, Department of the Prime Minister and Cabinet, viewed 1 October 2018, <http://pmtranscripts.pmc.gov.au/release/transcript-41039>.

Vincent, M 2019, 'Suspicion falls on China after cyber attack on Australian Parliament - and it's not surprising', ABC News*,* viewed 17 November 2019, <https://www.abc.net.au/news/2019-02-08/australian-parliament-cyber-security-breach-blame-on-china/10795010>.

Weiss, GW 1996, *The farewell dossier*, viewed 13 February 2020, <https://apps.dtic.mil/dtic/tr/fulltext/u2/a527328.pdf>.

Whitmore, P 2016, 'Current cyber wars', eds. LJ Janczewski & W Caelli, *Cyber Conflicts and Small States*, Ashgate, UK, pp. 47-69.

Williams, A 2019, 'A methodology for the comparative analysis of strategic culture and cyber warfare', *Proceedings of the 18th European Conference on Cyber Warfare and Security (ECCWS '20)*, University of Coimbra, Portugal, July, pp. 568-76.

Enhancing the European Cyber Threat Prevention Mechanism

J Simola

Laurea University of Applied Sciences
RDI Espoo, Finland
University of Jyväskylä, Finland

Email: simolajussi@gmail.com

Abstract: *This research will determine how it is possible to implement the national cyber threat prevention system into the EU level Early Warning System. Decision makers have recognized that lack of cooperation between EU member countries affects public safety at the international level. Separate operational functions and procedures between national cyber situation centres create challenges. One main problem is that the European Union does not have a common cyber ecosystem concerning intrusion detection systems for cyber threats. Also, privacy and citizens' security as topics are set against each other. The research will comprise a new database for the ECHO Early Warning System concept.*

Keywords: *Information Sharing, Cybersecurity, HAVARO, Privacy, Early Warning*

Introduction

This paper will comprise a new database for the ECHO (the European network of Cybersecurity centres and competence Hub for innovation and Operations) Early Warning System concept. E-EWS aims at delivering a security operations support tool which enables the members of the ECHO network to coordinate and share information in near real time. Within the E-EWS, partners of ECHO can retain their fully independent management of cyber-sensitive information and related data management. The Early Warning System will work as a parallel part of other mechanisms in the public safety environment. Crucial scientific literature, interviews, and official publications concerning cybersecurity information sharing generate fundamental knowledge to understand the main factors, which separate and combine EU member countries in this environment. The purpose is to support the technical designers of the E-EWS consortium to develop the Early Warning System. Also, interviews of the cybersecurity specialists form crucial sources for the paper.

The HAVARO, organized by TRAFICOM (the Finnish Transport and Communications Agency) and NESA (National Emergency Supply Agency), is one kind of national early warning system, which gathers threat-informed data and produces crucial information concerning the situation of cybersecurity information sharing within critical infrastructure (Ladid, Armin & Kivekäs 2019).

This paper will explore those factors (requirements) which affect the conversion of a national EWS to a common early warning ecosystem at the EU level. Every EU member country has its own system for monitoring and protecting the cyber domain among vital functions. It must be understood

ISSN 1445-3312 Print/ISSN 1445-3347 Online

that national systems must find common procedural and governance models in the name of the common good. In addition, privacy-issue-related problems concern the whole cyber ecosystem. The public safety sector will not operate in an isolated dimension without connection to private sector companies. The crucial question is how to combine and share relevant data between stakeholders at the national level and at the international level.

The paper starts with a section introducing the background of challenges concerning critical infrastructure protection and discusses cybersecurity information sharing at the EU level and with the U.S. The next section handles the national HAVARO system and system requirements. The paper concludes with suggestions for a bases of the solution and conclusions about the research area.

Challenges Concerning Critical Infrastructure Protection

According to the Horizon 2020 work program, disruption in the operation of EU member countries within critical infrastructure may result from hazards and physical or cyber-physical events (European Commission 2019).

Public safety authorities have noticed in Finland that protecting modern infrastructures and vital functions needs not only to protect physical operative functionalities and equipment; they also need the cyber-dimension in their daily routine. It is possible to integrate cyber-threat-informed functionalities of the computer emergency response teams and operative functions of the public safety organizations. These integrated systems are examples of Cyber Physical Systems (CPS) that integrate computing and communication capabilities with monitoring and control of entities in the physical world (Secretariat of the Security Committee 2019).

In the European Union, there has been a common will to enhance cooperation between public authorities. According to the European Council (2010), Europol collects and exchanges information and facilitates cooperation between law-enforcement authorities in their fight against cross-boarding organized crime and terrorism. Eurojust drives coordination and increases the effectiveness of judicial authorities. Frontex manages operational cooperation at the external borders. The EU operates as the Counterterrorism Coordinator. Several networks have also been established in the fields of training, drugs, crime prevention, corruption, and judicial cooperation in criminal matters (European Council 2010). Solutions are based on common recognition for information sharing and are designed to ease joint investigations and operations. Instruments based on mutual recognition include the European Arrest Warrant and provision for the freezing of assets (European Council 2010). The report is only 10 years old, and only two lines of text have been used to analyse cyber threats.

There are separate local situation centres for emerging situations and emergency response systems, and there are separate cyber-threat functions at the national and EU level. All work mainly without synergy. ICT development projects—for example MARISA, EUCISE, and RAPID—are European-Commission-funded projects that are producing better common situational awareness among EU member countries. The main limitation to implement the RAPID system is related to a lack of cooperation between the EU countries and real-time features of the mechanism. In addition, a lack of leadership causes problems in collaboration (Apuzzo 2019).

One crucial thing is still missing: combined cyber-physical functionalities (Simola & Rajamäki 2017). It is not enough that there are national computer emergency response teams, which only monitor Internet traffic. In the future, there is a growing need to use proactive or preventive functionalities among public safety organizations.

Information Sharing at the EU Level and a National Intrusion-Detection System

Shared (cyber) situational awareness is closely related to (cybersecurity) information exchange (Bolstad & Endsley 2000). Bolstad and Endsley (2000) define the development of shared Situational Awareness as consisting of these four factors:

- Shared SA requirements (degree to which team members realize which information is needed by other team members);
- Shared SA devices (communications);
- Shared SA mechanism (shared mental models); and
- Shared SA processes (effective team processes for sharing relevant information).

According to Munk (2018) information interoperability is the joint capability of different actors—such as persons, organizations, and groups—necessary to ensure the exchange and common understanding of the information needed for their success.

The central government of Finland is one of the most important administrative actors that needs correct environment-related cyber situational awareness. When something abnormal occurs, different ministries try to gather and to share the same data from the site of an accident. The common cybersecurity information-sharing procedure enables the government to react to new kinds of threats. There is a need to create a common early warning system with preventive functions. Service producers may be based on public organizations and private companies. One of the most important things is that governance responsibilities of the operational functions should be designated in the future.

In partnership with the National Emergency Supply Agency (NESA), TRAFICOM created the system called HAVARO 1.0 in 2011 (National Cybersecurity Center-FI [NCSC-FI] 2019). It is optional for every Finnish organization to join the system. The information on situation awareness provided by the system increases understanding of the organization´s own and the general state of information security. The system produces information, which makes it possible to alert other players about a detected threat and to develop better tools of detection. The participating organizations are responsible for the costs of equipment needed for their network.

The companies and public administration operators participate in the HAVARO operation voluntarily. The operation of the system is based on the information security threat identifiers coming from different sources. With the help of the identifiers, harmful traffic can be detected from the organization's network traffic. The NCSC-FI receives the information about the anomalies and analyses them. In case of an information security threat, the organization is warned. Based on the information from the HAVARO, the other operators can also be warned about the detected threat. That way, the system helps not only individual organizations, but also helps form a general view of information-security threats against Finnish information networks. TRAFICOM provides the GovHAVARO service for the state administration operators. It completes the information and cybersecurity threat detection of the state administration's Internet traffic. The main problem with HAVARO 1.0 concerns the monitoring ability (Lehto *et al.* 2018). It mainly monitors information-security incidents in Internet traffic (KPMG 2013). It is incapable of monitoring the communication of individual user behaviour.

In the future, it is not enough to monitor only the Internet traffic of companies. There should be a wider right to access the organizations' information systems and communication because the Internet of Things (IoT) is changing the way the Artificial Intelligence atmosphere is understood. When electrical and telecommunication cables are placed in the same pipeline, possibilities for vulnerabilities increase.

The HAVARO service is now under development. Instead of being a government service, HAVARO 2.0 will be jointly provided by commercial operators and the NCSC-FI. Some of the events will be processed and reported by information Security Operations Centres (SOC). The objective of the HAVARO 2.0 project is to create the trust network in which the members can exchange information among themselves better than they have before. The HAVARO 2.0 Early Warning System will consist of features of the existing 1.0 system with developed early-warning dimensions. Existing cyber-threat sensor systems need more specialized detection features. Increasing the cyber-threat atmosphere will force stakeholders to develop a better and more efficient system. Separate forensics methods, gathering logs, gathering information, reverse engineering, and analysing risks are not enough in the future. It is crucial to produce added value by combining different data sources and weak threat signals. HAVARO 2.0 will only be complementary to other cybersecurity services.

HAVARO 2.0 will include the GovHavaro feature (Lehto *et al.* 2018). That means that there will be a connection between public organizations and the HAVARO Early Warning System. This information is classified as more confidential, but sector-based sharing requires the sharing of this information to all public safety organizations and to the central government. At the EU level, this information is important to be shared in real time to the stakeholders if threat-information regarding cybersecurity related information to other countries or threat information generates a common risk to vital functions. New stakeholders of the HAVARO 2.0 have contractual relationships with SOCs, not with the NCSC.

Cybersecurity Information Sharing with the U.S.

There are no fundamental differences in administrative functions between the European Union and the United States. Mainly there are more similarities than differences. Legislation and regulation between the U.S. and the EU are coming closer to each other. The NIS directive in the EU will help to develop next-generation early warning systems.

According to the European Parliament and the Council of the European Union (2016), General Data Protection Regulation (GDPR) was designed to harmonize data-privacy laws across Europe, to protect and empower all EU citizens' data privacy, and to reshape the way organizations across the region approach data privacy. GDPR applies to all businesses offering goods and/or services to the EU. That means, if a company is holding private information about an EU citizen to whom it provides services, GDPR applies. It strengthens the rights of private information, access, and the right to be forgotten. The GDPR protects personal data regardless of the technology (automated and manual processing) used. GDPR concerns both unions. The U.S. and the EU have made fundamental agreements to generate a common base for fluent information sharing (European Parliament and the Council of The European Union 2016). Public safety actors, like European law enforcement agencies, need a common situational picture for the cross-boarding tasks so that operational cooperation will be based on a reliable platform.

The European Commission presented the cybersecurity strategy of the European Union in 2013. It set out the EU approach on how to best prevent and to respond to cyber disruptions and attacks as well as emphasized that fundamental rights, democracy, and the rule of law need to be protected in the cyber domain. Cyber resilience is one of the strategic priorities. That means that effective cooperation between public authorities and the private sector is a crucial factor, that the national Network and Information Sharing competent authorities should exchange relevant information with other regulatory bodies.

The information sharing between the EU and the U.S. has been regulated among other things, as follows; the European Commission and the U.S. Government reached a political agreement on a new framework for transatlantic exchanges of personal data for commercial purposes named the EU-US Privacy Shield (European Commission 2016). The framework protects the fundamental rights of anyone in the EU whose personal data is transferred to the United States as well as brings legal clarity for businesses relying on transatlantic data transfers. The EU-US Privacy Shield is based on several principles that govern companies that handle data. They are as follows: a) the U.S. Department of Commerce will conduct regular updates and reviews of participating companies to ensure that companies follow the rules they submitted themselves to; b) the U.S. has given the EU assurance that the access of public authorities for law enforcement and national security are subject to clear oversight mechanisms; c) citizens who think that collected data has been misused under the Privacy Shield scheme will benefit from several accessible dispute resolution mechanisms. It is possible for a company to resolve the complaint by itself or give it to the Alternative Dispute Resolution (ADR) to be resolved for free. Citizens can also go to their national Data Protection Authorities, who will work with the Federal Trade Commission to ensure that complaints by EU citizens are investigated and resolved. The Ombudsperson mechanism means that an independent senior official within the Department of State will ensure that complaints are properly investigated and addressed in a timely manner (European Commission 2016).

According to the U.S. Department of Commerce (2020), the United States has taken a different approach to improving the protection of privacy from that taken by the European Union. The United States uses a sectoral approach that is based on a combination of legislation, regulation, and self-regulation. The approach provides organizations in the United States with a reliable mechanism for personal data transfers to the United States from the European Union. This mechanism ensures that EU data subjects continue to benefit from effective safeguards and protection as required by European legislation with respect to the processing of their personal data when it has been shared to outside of the EU area. The Department of Commerce is issuing these Privacy Shield Principles, including the Supplemental Principles under its statutory authority to foster, promote, and develop international commerce (U.S. Department of Commerce 2020).

Challenges with the Privacy Shield Agreement

Privacy activists have challenged the Privacy Shield Agreement by arguing that U.S. national security laws did not protect EU citizens from government snooping. On 16 July 2020, the EU Court of Justice made the decision about the adequacy of the protection provided by the EU-US Data Protection Shield by invalidating the agreement (Court of Justice 2020). Despite this decision, the EU Commission Decision on standard contractual clauses for the transfer of personal data to processors established in third countries is valid. Affected companies will now have to sign ‘standard contractual clauses’—non-negotiable legal contracts drawn up by Europe, which are used in other countries besides the U.S. As regards the requirement of judicial protection, the Ombudsperson mechanism referred to in that decision does not provide data subjects with any cause of action before a body which offers guarantees substantially equivalent to those required by EU law, such as to ensure both the independence of the Ombudsperson provided for by that mechanism and the existence of rules empowering the Ombudsperson to adopt decisions that are binding on the U.S. intelligence services. For the above, the Court of Justice declared the European Commission Decision 2016/1250 invalid (Court of Justice 2020).

The purpose of standards is to simplify the work of authorities, to facilitate trade, and to make consumers' everyday lives easier. Standardization helps companies and enterprises to create common rules for information sharing and data handling. The family of 270XX standards provides the bases for the definition and implementation of an Information Security Management System (ISMS). For example, standard ISO/IEC 27010:2015 belongs to an ISO 27000 family and is a key component of trusted information sharing. This International Standard is applicable to all forms of exchange and sharing of sensitive information, both public and private, nationally and internationally, within the same industry or market sector or between sectors (International Organisation for Standardisation 27010:2015).

A trusted independent entity would be appointed by the information-sharing community to organise and to support their activities, for example, by providing a source anonymization service (International Organisation for Standardisation 27010:2015).

ISO standard 11179 (2019) provides guidelines for the naming and definition of data elements, as well as information about the metadata captured about data elements (International Organisation for Standardisation 11179-7:2019). Standard 24745 (2011) ensures that any information that identifies or can be used to identify, contact, or locate the person to whom such information pertains; from which identification or contact information of an individual person can be derived; or that is or might be directly or indirectly linked to a natural person be kept private. These are only examples of a wide range of standards that companies must follow. Standardization strengthens product compatibility and safety, protects the citizens, and protects the environment (International Organisation for Standardisation 24745:2011).

System Requirements

Humans are not as good at processing large volumes of data—quickly and consistently. Flexible autonomy should provide a smooth, simple, seamless transition of functions between the human and the system (Endsley 1988).

National early warning system and information sharing among ECHO EWS partners sets requirements for the basis of the research. Collected materials comes from the scientific literature, interviews of IT specialists, research articles, and official publications.

ECHO EWS will deliver a secure sharing support tool for public-safety personnel to coordinate and to share information in near real-time. It will support information sharing across organizational boundaries and will provide the sharing of general cyber information as a reference library. It will also ensure secure connection management from clients accessing the E-EWS. It will combine different kinds of functions required in the management of information-sharing functions, including sector-specific cyber-sensitive data. All participants (administrative actors, EU countries, companies, cyber situational centres, and public safety authorities) set requirements for developing

ECHO system governance and the Early Warning System. The big challenge is the diversity of stakeholders included in the ECHO. Therefore, system requirements cannot place too many challenging barriers to the development of the E-EWS.

When the aim is to share essential information between stakeholders as soon as possible, information sharing must be automatized. AIS (Automatic Identification System) utilizes the Structured Threat Information Expression (STIX) and Trusted Automated Exchange of Indicator Information (TAXII) specifications for machine-to-machine communication. STIX is a language and serialization format that enables organizations to exchange Cyber Threat Intelligence (CTI) in a consistent and machine-readable manner. Trusted Automated eXchange of Intelligence Information (TAXII™) is an application layer protocol used to exchange cyber threat intelligence (CTI) over the HTTPS (Department of Homeland Security [DHS] 2019). Echo EWS system requirements are based on requirements concerning governance model and Echo Federated Cyber Range.

Bromander, Muller and Jøsang (2020) have criticized the use of STIX because of various ways of representing the same information, the possibility of automatic consumption, and the fact that computer-based analysis becomes limited. If a computer cannot identify information because the information type is not normalized, 'Big Data'-style analysis is not possible; therefore, manual work is needed to correct and to analyse the data Also lack of standardization concerning all relevant information poses a problem for automation. Bromander, Muller and Jøsang (2020) argue that while many claim to use STIX, in most cases it is not used as a standardized way of sharing CTI suitable for automation. The criticism is justified and seems to concern large companies. However, there are currently no well-developed alternative good solutions.

Suggestion for a Basis of the Solution

This section describes the findings and suggested basis of a solution for national information sharing. First, the information-sharing architecture in the U.S. will be addressed. After that, methodologies for the indicator sharing and possible features for the early warning system will be introduced.

Information-sharing architecture in the U.S.

NCSC-FI (National Cybersecurity Center) and NESA (The National Emergency Supply Agency) have made an industry-specific classification for sharing cyber-threat information. The classification is demonstrated as follows: VIRT, public organizations, defense industry, energy sector, finance, industry automation, chemical and process industry, logistics sector, food industry, health sector, industrial companies, equipment and product manufacturers, ICT, media industry, security consultants, security researches, CERT-actors. Despite the classification, there is a need to expand collaboration within public and private actors. NESA, as a partner of TRAFICOM, is responsible for vital functions of society in Finland (NCSC-FI 2017). This classification mainly follows the European model, but also follows the sector-based classification in the U.S.

As mentioned above, the information-sharing model used in the U.S. is possible to replicate in the European Union. There are more similarities than differences. The simple picture in **Figure 1**, below, shows how information is shared. Automated information (indicator) sharing is mainly based on centralized ISACs, which consist of all actors of the specific sector. As illustrated in **Figure 1**, below, sector-based Information Sharing and Analysis Centers (ISACs) are one kind of government-prompted, industry-centric sharing model. Centers are non-profit, member-driven organizations formed by critical infrastructure owners and operators to share information between government and industry (ENISA & ITE 2017). Finland uses a similar national level structure of information sharing. It is based on the classification of different sectors of critical infrastructure. There are 16 levels of critical infrastructure used in the U.S. The same sector-specific frame is almost in use everywhere in western countries (White House 2013a; 2013b).

Open Communities and Platforms are open-source sharing platforms. For example, STIX indicators and open-source intelligence feeds are this kind of format. The Malware Information Sharing Platform (MISP) is a free, open-source platform developed by researchers from the Computer Incident Response Center of Luxemburg, the Belgian military, and NATO. For example, Interpol uses the Malware Information Sharing Platform (GitHub 2019; OASIS Cyber Threat Intelligence (CTI) TC, DHS (CS&C) 2017a).

HAVARO as a part of the European Early Warning System

There are several factors that are important to notice if the purpose is to integrate the national Early Warning System to the common European Union level Early Warning System. First, the use of cloud services is not a secure way to store and gather threat-informed data. When customers of the early warning solution are connected to the system from all around Europe, using cloud-only service solutions is not secure because cyberattacks against virtual machines may jam the whole system. Therefore, the authors recommend using a centralized main server that produces services to EWS stakeholders. This sharing model requires using local (national) E-EWS servers where ECHO-EWS is connected This is one kind of hybrid model, but the model is a secure part of the architecture, which allows sharing trust-level information. It is important that, for example, the National Bureau of Investigation have the ability to gather and to share trust-level information concerning vital functions of society and have the ability to be connected in the Early Warning System. It is relevant that the early warning data is shared from the central server to the affected sectors. International researchers recommend using a controlled information-sharing model, where national public safety actors share relevant data to stakeholders via a centralized center (EWS Center [Department of Homeland Security]) as **Figure 1** illustrates.

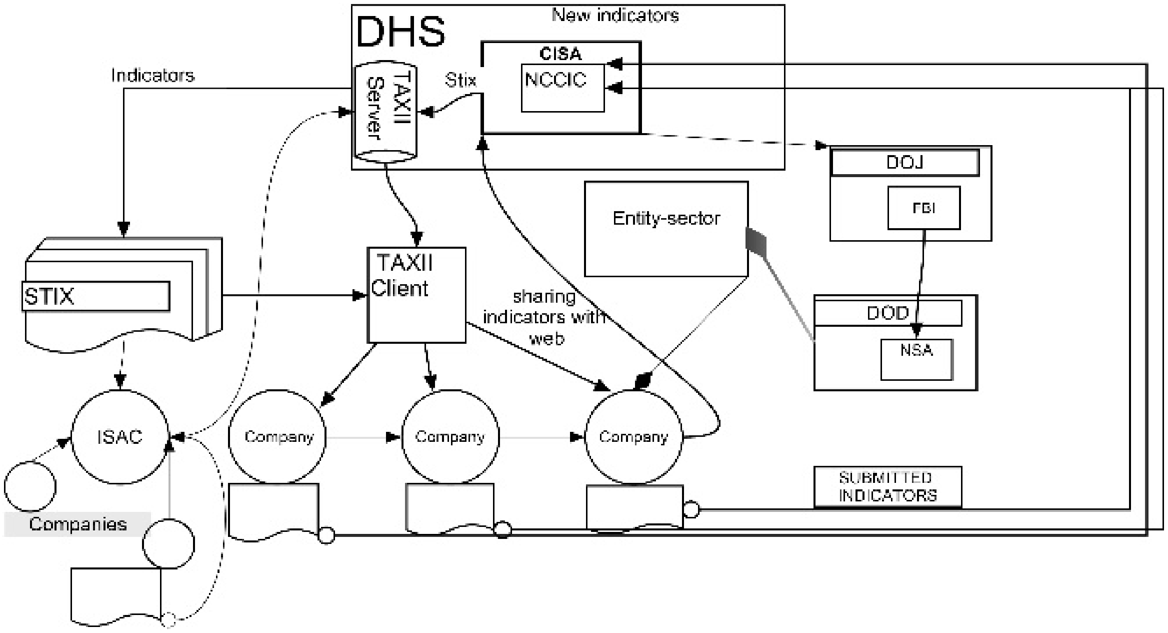


Figure 1: Cyber-information sharing model in the U.S.

Two-way models also allow public safety organizations to use gathered information for the prevention of hybrid threats before the domino effect is caused by two or more separate phenomena. It is important that cross-boarding cooperation work directly and instantly. Echo EWS will not work as a separate system but plays a crucial and parallel part in wider mechanisms, including the European-level situational awareness system of NATO. All Echo partners must understand that common language means in a wider manner—for example, taxonomies, techniques, procedures, and common ways to respond and act.

The U.S. Department of Homeland Security uses a system called Automated Indicator Sharing (AIS). AIS participants may connect to a national early warning system in the National Cybersecurity Center (NCSC) that allows also bidirectional sharing of cyber threat indicators. A server housed at each stakeholder´s (community) location allows the stakeholder to exchange indicators with the National Cybersecurity Center (NCCC) as **Figure 1** illustrates. Participants receive and can share DHS-developed indicators that they have observed in their own network defence efforts, which the national cyber situation centre will then share back out to all AIS participants. Stakeholders who share indicators through AIS will not be identified as the source of those indicators to other participants unless they consent to the disclosure of their identity. Senders are anonymous unless they want NCSC to share their identity (Hernandez-Ardieta, Tapiador & Suarez-Tangil 2013). Official cyber-security partners will vet the indicators they receive through AIS.

The government also needs useful information about indicators and other threat-informed data. Therefore, local NCSC should share at least weekly reports to the government situation centre. AIS utilizes the Structured Threat Information Expression (STIX) and Trusted Automated Exchange of Indicator Information (TAXII) specifications for machine-to-machine communication. STIX is a language and serialization format that enables organizations to exchange Cyber Threat Intelligence (CTI) in a consistent and machine-readable manner. Trusted Automated eXchange of Intelligence Information (TAXII™) is an application layer protocol used to exchange cyber threat intelligence (CTI) over the HTTPS (Department of Homeland Security 2019).

Collection-based communications indicate that a single TAXII client is making a request to a TAXII server and the TAXII Server carries out that request with information from a database. A TAXII channel in TAXII Server enables TAXII clients to exchange information with other TAXII clients in a publish-subscribe model. TAXII clients can push messages to Channels and Subscribe to Channels to receive published messages. A TAXII Server may host multiple channels per API root (MITRE 2018; OASIS Cyber Threat Intelligence [CTI] TC, DHS [CS&C] 2017b). TAXII is the main transport mechanism for Cyber Threat Information (CTI) represented in STIX. Stakeholders may share indicators with NCSC through an ISAC or an ISAO without being a TAXII client.

According to the Department of Homeland Security (2019) Cyber Threat Information is any information related to a threat that might help an organization protect itself against a threat or detect the activities of an actor.

There are a wide range of the information-sharing methodologies and systems in law enforcement. For example, the main approach of the Europol Information System (EIS) is to be the reference system for offenses, individuals involved, and other related data to support EU member states, Europol, and its cooperation partners in their fight against organized cybercrime, terrorism, and other forms of serious crime. For example, the European Cybercrime Centre (EC3), as a part of Europol, uses an open source based MISP platform (ENISA 2017). Malware Information Sharing Platform (MISP) is a tool for information sharing about malware samples and related malicious campaigns related to specific malware variants. It offers architectural flexibility and allows the use of a centralized platform (for example, CIRCL and FIRST instances), but also as a decentralized (peer-to-peer) platform.

Europol´s SIENA is a VPN (Virtual Private Network) designed to enable a swift, secure, and user-friendly exchange of operational and strategic crime-related information and intelligence between member states, Europol, law enforcement cooperation partners, and public safety organizations (EUROPOL 2019).

Databases of the Schengen Information System (SIS) and networks have also been established for the exchange of information on criminal records, on combating hooliganism, on missing persons or stolen vehicles, and on visas which have been issued or refused. DNA and fingerprint data help put a name to anonymous criminals who left crime scenes. EU legal instruments facilitate operational cooperation between member states, such as the setting up of collaborative investigation teams and the organizing of joint operations (European Council 2010).

Sharing digital information between stakeholders may include Common Vulnerabilities and Exposures (CVE) or CVE-ID and CVEs that include a list of common identifiers for publicly known cybersecurity vulnerabilities. For example, the HAVARO EWS solution exploits identifiers to detect threats. CVE Numbering Authorities (CNAs) are authorized organizations which assign CVE IDs to vulnerabilities affecting products within their distinct agreed-upon scope for inclusion in first-time public announcements of new vulnerabilities (MITRE Corporation 2019a). MITRE Corporation (2019b) CVE Identifiers are unique, common identifiers for publicly known information security vulnerabilities (MITRE Corporation 2019b).

The National Vulnerability Database (NVD) is the U.S. government repository of standards-based vulnerability management data represented using the Security Content Automation Protocol (SCAP). This data enables automation of vulnerability management. The NVD consists of databases of security checklist references, security-related software flaws, misconfigurations, product names, and impact metrics (NIST 2019).

In the CVE list feeds, NVD and CVE entries provide enhanced data for each entry—such as fix information, severity scores, and impact ratings. NVD also supplies advanced searching features (MITRE Corporation 2019a; 2019b).

Digital Forensics XML (DFXML) is an XML language. DFXML improves composability by providing a language for describing forensic processes (for example, cryptographic hashing), forensic work products (for example, the location of files on a hard drive), and metadata (for example, file names and timestamps) (Garfinkel 2012).

According to Garfinkel (2012), the Digital Forensics XML toolset is intended to represent the following types of forensic data:

- Metadata describing the source disk image, file, or other input information.

- Detailed information about the forensic tool that did the processing (for example, the program name, where the program was compiled, and linked libraries).
- The state of the computer on which the processing was performed (for example, the name of the computer, the time that the program was run, the dynamic libraries that were used).
- The evidence or information that was extracted (how it was extracted and where it was physically located); cryptographic hash values of specific byte sequences; operating-system-specific information useful for forensic analysis (Garfinkel 2012).

Conclusion

The fight against hybrid threats means not only preventing functions against cyberattacks, but also identifying, tracing, and prosecuting a criminal/criminal group. This means even multifunctional integration where existing intrusion detection/prevention systems complement new solutions in the future.

There are no essential barriers to increase collaboration in organizational, tactical, strategical, and technical levels between national CERTs, NATO Computer Incident Response Capability (NCIRC), and EU Computer Emergency Response Team (CERT-EU). Common E-EWS solution would create an effective way to respond to cross-boarding hybrid thread situations. All major companies whose businesses are involved with the vital functions of society should be connected to an early warning system.

The future HAVARO 2.0 that is under development reflects a tendency to develop early warning functions at the national level. However, this is not enough. Critical information must be able to share between EU member countries because several enterprises operate at the international level. Cross-border cyber threats force countries to exchange critical information within EU member countries and between EU and other western states. That means cyber risks have become common challenges.

Operative public safety functions require quicker response or even prediction. HAVARO 2.0 should utilize the Artificial Intelligence (AI) dimension to detect threats. It is not possible to design next-generation early warning information systems without machine learning as part of the Artificial Intelligence (AI) functionalities because the early warning system requires predictive features. Artificial Intelligence functionalities enable entities to exploit difference databases and produce characterized data more effectively than a human can; it may also come to a conclusion by learning from input information. In addition, AI can make a decision without human interaction. This means also that not every ECHO participant has the same potentiality or opportunity to develop national system architecture. International cyber-physical dimension of threats sets requirements, what should be the minimum cybersecurity level or requirements of cyber situational centers at the national level. Framework for the local, national, and international information sharing should follow the same principles in each EU member country. **Figure 2**, below, illustrates the simple formation of cybersecurity information sharing between countries in which HAVARO 2.0 may join. This example consists of separate national sub-hubs and one centralized hub. Information-Sharing participants do not exchange information with each other. All threat-informed data is shared via a hub.

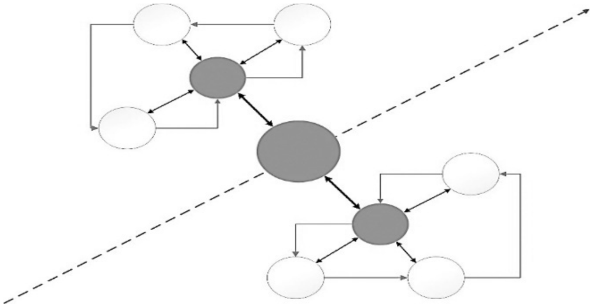


Figure 2: Connection between sub-hubs

Therefore, ISAC based national sectorial classification is the optimal way to share classified information as **Figure 3** illustrates.

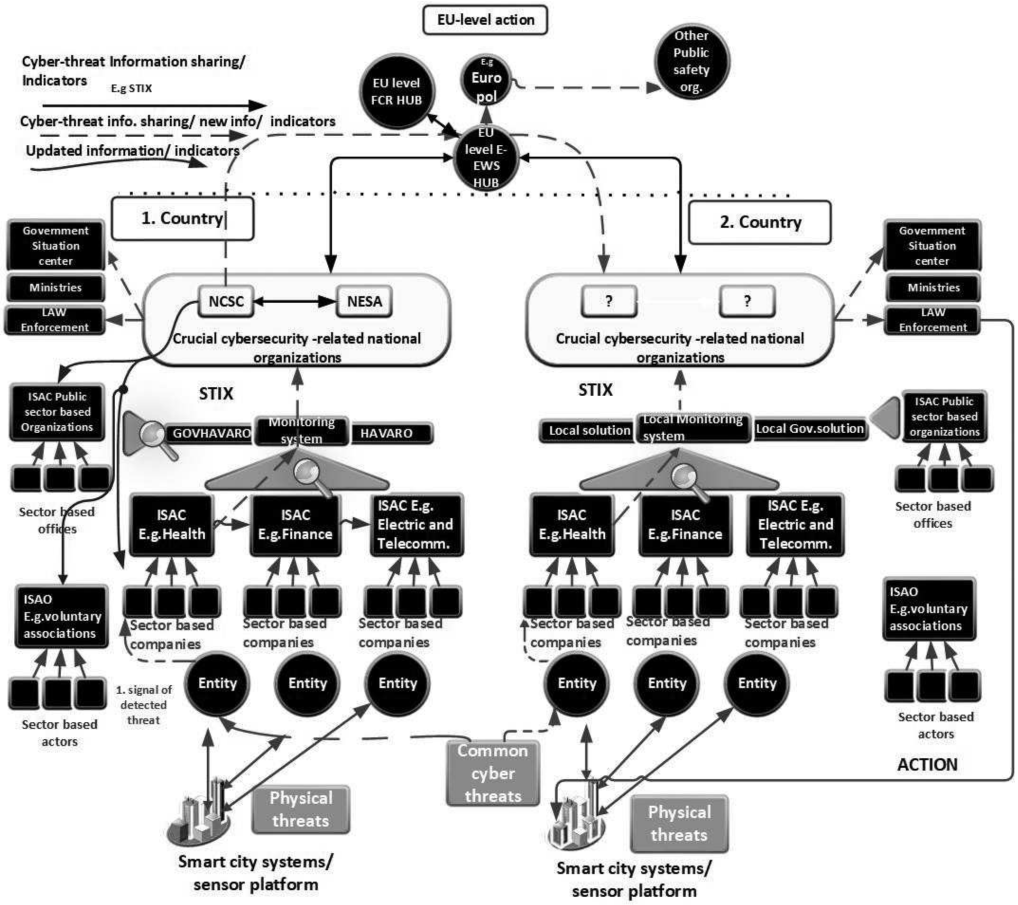


Figure 3: Proposed E-EWS information-sharing model

Figure 3 demonstrates information-sharing relationships and organizational structures concerning information sharing within a centralized hub system (countries, companies, public safety organizations, and other actors). In country number 1 (Finland), identifiers of the national Early Warning System (for example, HAVARO) detect a weak signal of cyberthreat concerning Internet traffic in a multinational enterprise. The national cybersecurity centre of country 2 has not noticed a cyberthreat activity. Automated Information Sharing functionalities produces crucial data for the central EWS hub, which shares relevant information in near real-time to the situation centres (CERT or

CIRT team). Sensitive data will be shared directly to the international public safety organizations and/or to the governments which are associated with the cyberthreat. NCSC of Finland uses a parallel subsystem for public organizations; HAVARO consists of separate early warnings solutions named “GovHavaro” for all public organizations.

Participants do not need to share information directly with each other, but there is a need to establish sector-specific communities—for example, ISAC and ISAO—that collect crucial information concerning the targeted sector of the critical infrastructure. This cybersecurity information is monitored and handled by national CERT or CIRT, and cybersecurity centres will share all new indicators between stakeholders (ISACs). All law enforcement-related information will be shared directly via EWS hub to the public safety authorities, such as EUROPOL or INTERPOL. Centralized EWS hub and sub-hubs are the simplest option for the national Finnish Early Warning System. On the other hand, a big challenge will be who maintains the central hub, and what its governance model would be.

Criticism concerning the use of STIX is justified, as mentioned above, and the problem needs to be rectified. More detailed guidelines, methods, standardization, and compliance with the law create a better operating environment to take advantage of automated indicator exchange.

Despite the invalidated privacy shield decision of the EU Court of Justice, there is a need to strengthen and to be aware of hybrid threats in a wider perspective. Privacy issues are important to protect. It is possible that the content of the privacy shield agreement needs to be changed. The agreement is significant in terms of commerce. Companies will now have to sign ‘standard contractual clauses’: non-negotiable legal contracts drawn up by Europe, which are used in other countries besides the U.S. (Court of Justice 2020).

References

Apuzzo, M 2019, ‘Europe built a system to fight Russian meddling. It is struggling’, *The New York Times*, viewed 1 November 2019, <https://www.nytimes.com/2019/07/06/world/europe/europe-russian-disinformation-propaganda-elections.html->.

Bolstad, C & Endsley, M 2000, ‘The effect of task load and shared displays on team situation awareness, *The 14th Triennial Congress of the International Ergonomics Association and the 44th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA, US.

Bromander S, Muller, EM & Jøsang A 2020, ‘Examining the “known truths” in cyber threat intelligence – The case of STIX’, *Proceedings of the 16th International Conference on Cyber Warfare and Security*, Old Dominion University, Norfolk, VA, US, pp. 493-502.

Court of Justice of the European Union 2020, ‘The Court of Justice invalidates decision 2016/1250 on the adequacy of the protection provided by the EU-US Data Protection Shield’, Press release No 91/20, 16 July , viewed 1 June 2020 <https://curia.europa.eu/jcms/jcms/p1_3117870/en/>.

Department of Homeland Security (DHS) 2019, ‘Automated Indicator Sharing (AIS)’, viewed 1 June 2019, <https://www.us-cert.gov/ais>.

Endsley, MR 1988, ‘Design and evaluation for situation awareness enhancement’, *Proceedings of the Human Factors Society 32nd Annual Meeting*, pp. 97-101.

ENISA 2017, ‘Tools and methodologies to support cooperation between CSIRTs and law enforcement version 1.0’ November, Heraklion, GR,

——& ITE 2017, ‘Information sharing and analysis centres (ISACs) cooperative models’, Heraklion, GR.

European Commission 2013, ‘Joint Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions,’ viewed 3 June 2020, Brussels, BE, < https://eur-lex.europa.eu/procedure/EN/2023 69>.

——2016, ‘EU-U.S. Privacy Shield: Stronger protection for transatlantic data flows’, Brussels, BE.

——2019, ‘14. Secure societies: Protecting freedom and security of Europe and its citizens’, *Horizon 2020 - Work Programme 2018-2020*.

European Council 2010, ‘Internal security strategy for the European Union towards a European security model’, General Secretariat of the Council, European Union, Brussels, BE.

European Parliament and the Council of The European Union 2016, ‘Regulation (EU) 2016/679 of the of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General Data Protection Regulation)’, *Official Journal* L 119, 4 May, viewed 1 August 2019, <https://eurlex.europa.eu/eli/reg/2016/679/oj>.

EUROPOL 2019, ‘Secure Information Exchange Network Application (SIENA)’. viewed 1 August 2019, <https://www.europol.europa.eu/activities-services/services-support/information-exchange/secure-information-exchange-network-application-siena>.

Garfinkel, S 2012, ‘Digital forensics XML and the DFXML toolset’, *Digital Investigation*, vol. 8, pp. 161-74.

GitHub 2019, ‘Support your workflow with lightweight tools and features’, viewed 7 July 2019, <https://github.com/MISP/MISP-Taxii-Server>.

Hernandez-Ardieta, JL, Tapiador, JE & Suarez-Tangil, G 2013, ‘Information sharing models to cooperative cyber defence’, *Proceedings of the 5th IEEE International Conference on Cyber Conflict (CyCon) 2013*, pp. 1-28.

International Organization for Standardization 2011, ‘Information technology — Security techniques — Biometric information protection ISO/IEC 24745:2011’, viewed 5 July 2020, <https://www.iso.org/standard/52946.html>.

——2015, ‘Security techniques information security management for inter-sector and inter-organizational communications’, ISO/IEC 27010:2015, viewed 5 July 2020, <https://www.iso.org/standard/68427.html>.

——2019, ‘Metadata registries (MDR) — Part 7: Metamodel for data set registration’, ISO/IEC 11179-7:2019, viewed 5 July 2020, < https://www.iso.org/standard/68766.html>.

KPMG 2013, ’IDS:N käyttöönotto herättää todellisuuteen’, viewed 5 July 2019, <https://www.hackingthroughcomplexity.fi/2013/04/idsn-kayttoonotto-herattaa.html>.

Ladid, L, Armin, J & Kivekäs H 2019, ‘The Finish electronic communications regulator TRAFICOM - A cybersecurity reference model for Europe’, SAINT Consortium/ TRAFICOM, Helsinki, FI.

Lehto, M, Limnéll, J, Kokkomäki, T, Pöyhönen, J & Salminen, M 2018, ‘*Kyberturvallisuuden strateginen johtaminen Suomessa* No. 28’, *Valtioneuvoston kanslia*, Helsinki, FI.

MITRE Corporation 2018, ‘Trusted Automated eXchange of Indicator Information - TAXII™ enabling cyber threat information exchange’, U.S Government.

——2019a, ‘Common vulnerabilities and exposures’, viewed 6 July 2020, <https://cve.mitre.org/cve/cna.html>.

——2019b, ‘CVE-details’, viewed 6 June 2020, <https://www.cvedetails.com/cve-help.php>.

Munk, S 2018, ‘Interoperability services supporting information exchange between cybersecurity organisations’, *Academic and Applied Research in Military and Public Management Science*, vol. 17, no. 3, pp. 131-48.

National Cybersecurity Center-Finland (NCSC-FI) 2017, ‘*Viestintäviraston kyberturvallisuuskeskuksen palvelut*’, Brochure Cybersecurity services of the NCSC-FI. Helsinki: TRAFICOM.

——2019, ‘Havaro service and FAQ’, viewed 5 July 2020, <https://www.kyberturvallisuuskeskus.fi/en/havaro-service>.

NIST 2019, ‘National vulnerability database - General information’, viewed 1 September 2019, <https://nvd.nist.gov/general>.

OASIS Cyber Threat Intelligence (CTI) TC, DHS (CS&C) 2017a, ‘STIX™ version 2.0. Part 2: STIX objects No. stix-v2.0-wd03-part2-stix-objects) OASIS open’.

——2017b, TAXII™ version 2.0. ‘Committee specification 01 No. taxii-v2.0-cs01) OASIS Open’.

Secretariat of the Security Committee 2019, ‘Finland´s cybersecurity strategy - Government resolution’, Ministry of Defense, Helsinki, FI.

Simola, J & Rajamäki, J 2017, 'Hybrid emergency response model: Improving cyber situational awareness', *Proceedings of the 16th European Conference on Cyber Warfare and Security*, University, College, Dublin, IE, pp. 442-51.

United States Department of Commerce 2020, 'The Privacy Shield framework in the United States', viewed 6 July 2020, <https://www.privacyshield.gov/Privacy-Shield-Principles-Full-Text>.

White House 2013a, 'Critical infrastructure security and resilience', Presidential Policy Directive, USC.

——2013b, 'Federal register - Improving critical infrastructure cybersecurity, Part III - Executive Order 1363', vol. 77, USC.

Mini-Drone Swarms: Their Issues and Potential in Conflict Situations

M Lehto[1], W Hutchinson[2]

[1] *University of Jyväskylä*
Finland

Email: martti.j.lehto@jyu.fi

[2]*Security Research Institute*
Edith Cowan University
Perth, Australia

Email: w.hutchinson@ecu.edu.au

Abstract: *Drones are currently used for a wide range of operations, such as border surveillance, general surveillance, reconnaissance, transport, aerial photography, traffic control, earth observation, communications, broadcasting, and armed attacks.*

This paper examines the swarming and associated abilities to overwhelm a combatant as well as bring extra functionality by means of extra sensors spread throughout the swarm. The strategy of stealth is becoming increasingly less effective. Combatants can not only sense them, but can also successfully destroy them (although this cannot be said for nano-drones). For mini-drones, objectives can be enhanced by the strategy of overwhelming.

Keywords: *Drone, Security, Artificial Intelligence, Swarming, Surveillance, Suicide Drones, Networks, Autonomous Drones, Lethal Autonomous Weapons (LAWs)*

Introduction

There is no one standard when it comes to the classification of Unmanned Aircraft Systems (UASs), sometimes called Unmanned Aerial Vehicles (UAVs). Defence agencies have their own standard, and civilians have their ever-evolving loose categories for UASs. People classify them by size, range, and endurance and use a tier system that is employed by the military. A UAS is a "system whose components include the necessary equipment, network, and personnel to control an unmanned aircraft" (Department of Defense [DoD] 2014). In some cases, the UAS includes a launching element (DoD 2014).

Papireddy defines Unmanned Aircraft as a powerful system, that does not carry a human operator, that uses aerodynamic forces to provide vehicle lift, that can fly autonomously or be piloted remotely, that can be expendable or recoverable, and that can carry a payload (2015). The International Civil Aviation Organization (ICAO) employs the acronym RPAS (Remotely Piloted Aircraft System) for "A remotely piloted aircraft, its associated remote pilot station(s), the required command and

ISSN 1445-3312 Print/ISSN 1445-3347 Online

control links and any other components as specified in the type design" (ICAO 2019). As the world pioneer in the creation and implementation of regulations for the use of commercial Unmanned Aerial Vehicles, the French Directorate for Civil Aviation (DGAC) sees commercial UAVs as drones. In a general way, many countries use the term 'drone'. For many, UAV is used mostly in a military context, so 'drone' covers both civil and military (Altigator Unmanned Systems 2019).

This article uses the term 'drone' to cover the whole spectrum of aerial unmanned vehicles. It should be noted that this paper, concentrates on the aerial environment. However, drones have been produced for other environments, such as underwater, terrestrial, maritime, and space, as well as various environments hostile to humans (for instance, radioactive, chemically affected, and highly infectious areas). These environments will be brought up in the latter part of the paper where general issues surrounding drones will be discussed. The concept of swarming, which has changed the nature of drone use and effectiveness, is also highlighted.

Categories

Unmanned Aircraft Systems (UASs) can be roughly divided into fixed and rotary wings. Multi-rotor helicopters are referred to as multi-copters. Other classification elements include size, Maximum Gross Takeoff Weight (MGTW), range, and endurance. For combat, there are two main groups: Unmanned Combat Aerial Vehicles (UCAVs) and Unmanned Combat Aerial Rotorcrafts (UCARs). These can be categorized by performance and combat mission.

Multi-rotor multi-copters powered by an electric power source are manufactured with various numbers of engines. Most used are

- Quadrocopter (4 propellers, vertically oriented),
- Hexacopter (6 propellers, 6-angle, symmetrically mounted),
- Oktocopter (8 propellers, either 4 or 8 angles symmetrically mounted, in 4-angle installation, with the motors arranged in pairs on top of each other).

Table 1, below, illustrates classifications according to the U.S. Department of Defense (DoD) (Pennsylvania State University 2019).

Category	Size	Maximum Gross Takeoff Weight (MGTW) (kg)	Normal Operating Altitude (ft)	Airspeed (knots)
Small UAV Mini, Micro, Nano UAV	Length 15 cm - 2 m Nano UAVs can also be smaller	0-9	<1,200 ft Above Ground Level, AGL	<100
Medium UAV	5-10 m	9-25	<3,500 AGL	<250

Large UAV	> 10 m	<600	<18,000 Mean Sea Level	<250
Larger UAV	> 10 m	>600	<18,000 MSL	Any airspeed
Largest UAV	> 10 m	>600	>18,000 MSL	Any airspeed

Table 1: UAV classification according to the U.S. Department of Defense

Table 2, below, illustrates classification according to range and operating time (Pennsylvania State University 2019).

Category	Range (km)	Operating time
Very low close-range UAV	5	20-45 min
Close range UAV	50	1-6 hours
Short range UAV	> 150	8-12 hours
Mid-range UAV	< 1000	12-24 hours
Endurance UAV	> 10 000	24-36 hours

Table 2: UAVs classification according to range and operating time

In the late 1990s, the U.S. Armed Forces produced a classification according to the information of the UAV system provided to different user levels. This classification is shown in **Table 3**.

UAV	Capability
Micro Unmanned Aerial Vehicle (MUAV)	Producing information within a radius of less than 100 kilometres from its land station.
Tactical Unmanned Aerial Vehicle (TUAV)	Producing information within a radius of about 200 kilometres of its land station.
Medium Altitude Endurance Unmanned Aerial Vehicle (MAE)	Producing information within a radius of about 750 kilometres of its land station.
High Altitude Endurance Unmanned Aerial Vehicle (HAE)	Producing information for long-term and near-real-time information for the control of large areas.

Table 3: UAV clasification based capability

One group consists of UAVs which are focused on combat:

- UCAV, Unmanned Combat Aerial Vehicle;
- UCAR, Unmanned Combat Aerial Rotorcraft

UACV	Performance	Combat mission
Deep Penetration RPAS	Designed for full electromagnetic stealth	Designated to conduct reconnaissance and air strikes deep inside enemy territory
Combat RPAS	Designed for high G-forces and manoeuvrability	Designated to conduct air-to-air and air-to-ground combat in non-permissive and hostile air environments
Swarm RPAS	Forming a swarm	Designed for expendability and operating in large numbers
Carrier RPAS	Designed to carry an immense stock of long-range	Precision-guided air-to-air and air-to ground munitions, designed to project military power like naval aircraft carriers

Table 4: UCAV classification based on combat missions

Over the past two decades, Remotely Piloted Aircraft Systems (RPASs) have been fielded in increasing numbers across many nations and military services. It is very unlikely there will be a 'one-size-fits-all' solution for RPAS operations in a contested environment. In addition, Reconnaissance RPAS are expected to be upgraded and to continue the role of current Medium-Altitude Long-Endurance (MALE)/ High-Altitude Long Endurance (HALE) systems (JAPCC 2014).

Table 4, above, illustrates UCAV classification based on combat missions (Joint Air Power Competence Centre [JAPCC] 2014).

Drone Autonomy

Autonomy allows the reduction of the frequency at which the operators must interact with the drone supporting the implementation of more robust system solutions, where the role of the operators is to manage and to supervise, through appropriate human machine interface, the command and control functions without direct interaction.

There are various ways to discuss autonomy in weapon systems. Although precise definitions are critical for design and engineering purposes, understanding the debate about autonomy requires an acknowledgement of these differing uses of the term, typically centred on ethically relevant subprocesses of the system as a whole: targeting, goal-seeking, and the initiation of lethality (Payne 2017).

According to the U.S. DoD (2018), 'autonomy' is defined as the ability of an entity to independently develop and select among different courses of action to achieve goals based on the entity's knowledge and understanding of the world, itself, and the situation. Autonomous systems are governed by broad rules that allow the system to deviate from the baseline. This contrasts with automatic systems, which are governed by prescriptive rules that allow for no deviations. While early robots generally only exhibited automatic capabilities, advances in Artificial-Intelligence (AI) and Machine-Learning (ML) technology have allowed systems with greater levels of autonomous capabilities to be developed. The future of unmanned systems will stretch across the broad spectrum of autonomy, from remote-controlled and automated systems to fully autonomous ones.

Autonomous categories are

> **Human-in-the-loop:** In this mode, humans retain control of selected functions preventing actions by the AI without authorization; humans are integral to the system's control loop.
> **Human-on-the-loop:** The AI controls all aspects of its operations, but humans monitor the operations and can intervene when, and if, necessary.
> **Human-out-of-the-loop:** The AI-algorithms control all aspects of system operation without human guidance or intervention. The autonomous drone engages without direct human authorization or notification.

Autonomy results from delegation of a decision to an authorized entity to act within specific boundaries. An important distinction is that systems governed by prescriptive rules that permit no deviations are automatic, but they are not autonomous. The Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (2014) states that to be autonomous, a system must have the capability to independently compose and select among different courses of action to accomplish goals based on its knowledge and understanding of the world, itself, and the situation. 'Automatic' really means the drone can function alone but only by obeying a set of pre-set rules in response to sensors' inputs.

Drone Military Operations

The development of Unmanned Aerial Vehicles is intensifying as technology becomes cheaper. Drones can be used in a flexible manner in different tasks, such as intelligence, surveillance, target acquisition, and recognition missions, in strikes against surface targets, over-the-horizon relaying of information, Electronic Warfare (EW), Combat Search and Rescue (CSAR), Chemical, Biological, Radiological, and Nuclear Warfare (CBRN), logistic replenishments, and Counter Improvised Explosive Devices (C-IED) in a favourable environment or in areas where the risk level is elevated. In conflict situations, their functions can be changed to a multitude of purposes.

Drones are presumed to provide their services at any time, to be reliable, automatic, and often autonomous. Based on these assumptions, governmental and military leaders expect drones to improve national security through surveillance and/or combat missions. To fulfill their missions, drones need to collect and process data. Therefore, drones may store a wide range of information from troop movements to environmental data and strategic operations. The amount and kind of information needed make drones an extremely interesting target for espionage and, hence, endanger drones through theft, manipulation, and attacks.

Various types of air domination systems are being considered to enable a military force to dominate an area from the air for extended periods and to deny enemy movements and manoeuvring. Unmanned combat aircraft, these purposes can be divided into two categories according to their operating model: loitering or swarming.

In the U.S., current systems under consideration are standard weaponized drones or small expendable loitering weapons fitted with imaging sensors, such as the Low-Cost Autonomous Attack System (LOCAAS). Operating in swarms of 'intelligent munitions' weapons, the LOCAAS can autonomously search for and destroy critical targets, while aiming over a wide combat area.
A loitering weaponized drone (also known as a suicide drone or kamikaze drone) is a weapon

system category in which the weaponized drone or munitions loiters around the target area for some time, searches for targets, and attacks once a target is located. Loitering systems enable faster reaction times against concealed or hidden targets that emerge for short periods without placing high-value platforms close to the target area and allow more selective targeting as the actual attack mission can be aborted.

Drone Civilian Operations

Various UAVs are increasingly being used for various civilian purposes, such as government missions (law enforcement, border security, coastguard), firefighting, surveillance of oil and gas industry infrastructure, and electricity grids/distribution networks, traffic control, disaster management, agriculture, forestry and fisheries, civil engineering, earth observation and remote sensing, and communications and broadcasting. PricewaterhouseCoopers (PwC 2016) estimated the value added to the economy by drones at $127 billion. According to Single European Sky ATM Research (SESAR 2016), the growing drone marketplace shows significant potential, with European demand suggestive of a valuation in excess of EUR 10 billion annually, in nominal terms, by 2035 and over EUR 15 billion annually by 2050.

The development of the civil drone industry is dependent on the ability of drones to operate in various areas of the airspace, especially at very low levels. According to SESAR (2016), "In aggregate, some 7 million consumer leisure drones are expected to be operating across Europe and a fleet of 400,000 is expected to be used for commercial and government missions by 2050".

Critical infrastructure (CI) includes a large variety of elements from nuclear reactors, chemical facilities, water systems, logistics, and airports to healthcare and communications, and now drones are growing a very important part in this critical infrastructure environment. They have numerous tasks in critical infrastructure maintenance and protection. Human work is reduced, and tasks can be performed cost-effectively.

At the same time, CI must deal with the new and emerging threat of drones. The most headline-grabbing risks tend to be those of physical and electronic attacks. For example, drones could carry explosives into a nuclear power plant or get close enough to execute cyberattacks, causing disruptions or mechanical failures or even stealing sensitive data. The low-cost, global proliferation and capabilities of drones weighing less than 20 pounds make them worthy of specific focus. Future adversaries could use these small systems to play havoc with critical infrastructure both in the air and on the ground—thus, necessitating new actions to defend CI assets.

In organized civil disturbances, the availability of mini-drones can be used for surveillance by both law enforcement and those causing the disturbance. Swarming these drones (discussed in the next section) can give a lot of coverage for either side, and a myriad of sensors can provide various data. Also, in disaster events (such as fires, large crashes of many types, or epidemics) they can be used for intelligence and for delivery of such things as drugs; in fact, the variety of functions can be left to the imagination.

Drone Swarming

Various types of air domination systems are being considered to enable a military force to dominate an area from the air (and in the sea, on ground, and in space for that matter) for extended periods and to deny enemy movements and manoeuvring.

'Swarming' is the coordinated use of various drones which might be of different types, 'intelligence', size, and capabilities so they can act in unison. This use of swarming techniques (where numerous drones are used for one purpose) is of increasing interest. The decreasing cost of smaller drones (Hambling 2015) plus the built-in redundancy of swarms make the use of many drones for an attack much more appealing.

Current systems under consideration are standard weaponized UASs or small expendable loitering weapons, fitted with imaging sensors, such as the Low-Cost Autonomous Attack System (LOCAAS). Operating in swarms of 'intelligent munitions' weapons, the LOCAAS can autonomously search for and destroy critical mobile targets while aiming over a wide combat area (DoD 2014). Along with sensor autonomy, swarming drones will require the ability to self-navigate and self-position to collect imagery and signals efficiently (DoD 2005).

A loitering munition (also known as a suicide drone or kamikaze drone) is a weapon system category in which the munition loiters around the target area for some time, searches for targets, and attacks once a target is located. Loitering munitions enable faster reaction times against concealed or hidden targets that emerge for short periods without placing high-value platforms close to the target area, and allow more selective targeting as the actual attack mission can be aborted. Loitering munitions fit in the niche between cruise missiles and Unmanned Combat Aerial Vehicles, sharing characteristics with both. They differ from cruise missiles in that they are designed to loiter for a relatively long time around the target area, and from UCAVs in that a loitering munition is intended to be expended in an attack and has a built-in warhead.

Drones are currently in widespread use around the world, but the ability to employ a swarm of these systems to operate collaboratively to achieve a common goal will be of great benefit to national defence. A swarm could support lower operating costs, greater system efficiency, as well as increased resilience in many areas.

Drone swarms carry additional communications needs. Effective distributed operations require a battlefield network for drone-to-drone communications to allocate sensor targets and priorities and to position aircraft where needed. While the constellation of sensors and aircraft needs to be visible to operators, human oversight of many drones operating in combat must be reduced to the minimum necessary to prosecute the electronic warfare. Automated target acquisition will transfer initiative to the autonomous drone, and a robust, anti-jam communications network that protects against hostile jamming, capturing, and manipulation of data is a crucial enabler of drone swarming (DoD 2005).

Kallenborn (2018), from the U.S. National Defense University, defines 'drone swarm technology' as the ability of drones to autonomously make decisions based on shared information. This has the potential to revolutionize the dynamics of conflict. In fact, swarms will have significant applications to almost every area of national and homeland security. Swarms of drones could search the oceans for adversary submarines. Drones could disperse over large areas to identify and to eliminate hostile surface-to-air missiles and other air defenses. Drone swarms could potentially even serve as novel missile defences, by blocking incoming hypersonic missiles. On the homeland security front, security swarms equipped with chemical, biological, radiological, and nuclear (CBRN) detectors, facial recognition, anti-drone weapons, and other capabilities offer defences against a range of threats.

McMullan (2019) argues that swarming drones come in different shapes and sizes. For example, the U.S. Defense Advanced Research Projects Agency (DARPA) has been working on a program dubbed Gremlins—micro-drones the size and shape of missiles—designed to be dropped from planes and to perform reconnaissance over vast areas. On the other side of the spectrum is the larger XQ-58 Valkyrie drone (8.8 m in length).

A San Diego company, Kratos Defense & Security Solutions produces two classes of jet-powered autonomous drones, the UTAP-22 Mako and the XQ-58 Valkyrie, which would collaborate with manned fighter jets as a 'loyal wingman' for a human pilot. They can carry precision-guided bombs and surveillance equipment (Gregg 2019b).

In 2016, DARPA launched the OFFensive Swarm-Enabled Tactics (OFFSET) program that envisions future small-unit infantry forces using swarms comprising upwards of 250 small Unmanned Aircraft Systems (UASs) and/or small Unmanned Ground Systems (UGSs) to accomplish diverse missions in complex urban environments. By leveraging and combining emerging technologies in swarm autonomy and human-swarm teaming, the program seeks to enable rapid development and deployment of breakthrough capabilities. OFFSET aims to provide the tools to quickly generate swarm tactics, to evaluate those swarm tactics for effectiveness, and to integrate the best swarm tactics into field operations. OFFSET will develop an active-swarm tactics-development ecosystem and supporting open-systems architecture, including

1) An advanced human-swarm interface to enable users to monitor and to direct, potentially, a real-time, networked virtual environment that would support a physics-based, swarm tactics game and
2) A community-driven swarm tactics exchange (Chung 2016).

OFFSET program, by leveraging and combining emerging technologies in swarm autonomy and human-swarm teaming, seeks to enable rapid development and deployment of breakthrough capabilities. The program consists of five research and experiment areas: swarm technology, human-swarm teaming, swarm perception, swarm networking, and swarm logistics. **Figure 1**, below, illustrates the autonomous swarm capability development of the OFFSETprogram (Peters 2019).

In August 2019, DARPA tested OFFSET by using a swarm of autonomous drones and ground robots to assist with military missions. DARPA showed how its robots analysed two city blocks to find, surround, and secure a mock city building (Peters 2019).

Finland's Ministry of Defense (Finland MoD 2015) addresses the reality that, in some cases, drones can carry out missions better and cheaper than manned aircraft. The widespread proliferation of Micro Air Vehicles (MAV), which are difficult to detect, is on the cusp of becoming extremely challenging for air defences. Even the smallest drones are suitable for intelligence and Precision Guided Munitions (PGM) for target designation. Moreover, they can double as weapons, even inside buildings. The most radical concepts focus on replacing the intelligence-targeting fire chain; they aim at achieving a rapid weapons effect with the coordinated use of swarming Unmanned Aerial Vehicles. This requires sufficient survivability and cost-effectiveness from drones to saturate the defence.

In an article presented at the *14th Annual International Conference on Cyber Warfare and Security*

(*ICCWS 2019*), Haberl and Huemer (2019) described the drone swarm attack. In 2018, the Russian Ministry of Defence announced that 13 drones, which had been fitted with small bombs, managed to attack Russian bases in Syria. Such drones, which are intended to explode on impact, need to be modified in order to carry explosives. It is easy to imagine how 3D-printing could come in handy in this regard, especially since drones are capable of evading missile warning systems without any additionally needed infrastructure or equipment.

As defined below, swarming is the coordinated use of various drones which might be of different types, 'intelligence', size, and capabilities so they can act in unison. This use of swarming techniques, where numerous drones are used for one purpose is of increasing interest. Generally, when dealing with security-related actions, there are two main emphases—overwhelming force and deception. The decreasing cost of smaller drones (Hambling 2015) plus the built-in redundancy of swarms make the use of many drones for an attack much more appealing as they tend to overwhelm any countermeasures against them. Also, it can make deception much more difficult for the number of drones, but some drones that are disabled will still leave others to carry out the mission. Thus, at its simplest, an attack of sacrificial, impact aerial drones in a swarm makes an effective tactic which can overwhelm the opponent.

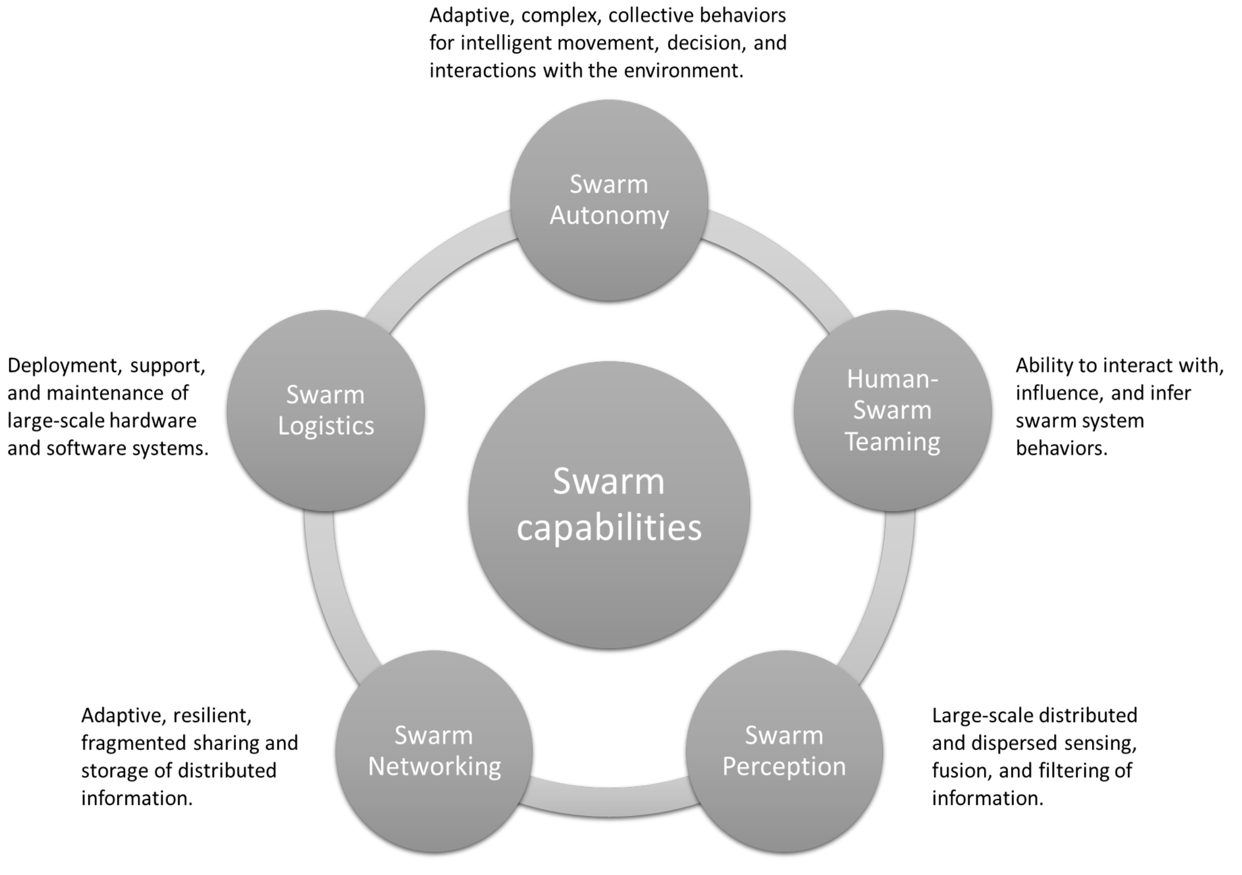

Adaptive, complex, collective behaviors for intelligent movement, decision, and interactions with the environment.
Swarm Autonomy
Deployment, support, and maintenance of large-scale hardware and software systems.
Swarm Logistics
Human-Swarm Teaming
Ability to interact with, influence, and infer swarm system behaviors.
Swarm capabilities
Adaptive, resilient, fragmented sharing and storage of distributed information.
Swarm Networking
Swarm Perception
Large-scale distributed and dispersed sensing, fusion, and filtering of information.

Figure 1: Autonomous swarm capability development in OFFSET program (modified from Chung 2016)

At present, these 'swarming drone systems' seem to be considered for underwater protection of valuable assets, such as submarines, and in the air for protecting manned aircraft, thus, providing surveillance for military units at a cheap cost. The initial use of 'tethered' drones linked to a 'mothership' gives protection to the controlling vehicle and its crew which, in turn, gives extra surveillance facilities and possibly firepower as well as cover by providing sacrificial drones to the central-control function. This concept develops into autonomous swarms, whereby each drone is independent but keeps communications with other drones and acts like one entity much like a flock of birds (Singer 2009). This implementation gives the group a lot more power and is much more difficult to deceive unless its elements are consistently very simple. However, simple, self-organizing swarms can lose some members without losing too much functionality, so deceiving and/or destroying the swarm will be harder than deceiving the individual. Nevertheless, swarms, because they need to link up with each other, are more vulnerable to infection from malware. Ironically, this vulnerability could be a weak point where software can be inserted.

Underwater drones do have a communication problem especially when not tethered to a 'mother ship' as communication signals are attenuated by the water medium. Signals are sent by radio and acoustic means or by light (blue has been used up to now). However, this has been partially overcome by using each drone arranged into a network of lines, passing the signal from one to another, and, thereby, extending the range much as classical network technology does.

The concept of swarms came out of a need to find asymmetric approaches to developing terrorist and insurgent approaches to war. From the U.S. perspective, the enemy in the early 21st century tended to be relatively small dispersed groups compared to conventional forces. Although these tactics were not new (Arquilla & Ronfeldt 2000), they did seem to be needed to compensate for the large, hierarchical forces which did not prove as flexible and speedy as these small groups. With the development and continuing advancement of military drones came the technological ability to produce smaller and more flexible varieties. As this development advanced, the increased communications and AI techniques allowed an ever-increasing potential of these machines to provide advanced drone swarms. The extension of network theory allowed the development of intelligent swarms which broadly can be hierarchical or networked (in an organizational sense).

Swarms can be designed so that the development of swarming systems can allow each element to work independently and come together in a swarm when needed so groups of drones can be expanded or decreased as the problem being tackled varies. Hence, drones of various abilities, functions, and forms can be coordinated, as necessary. This ability is very powerful and would require an opponent to work at a population level rather than targeting an individual drone. A well-chosen targeted drone might have the desired impact, but this would depend on the architecture of the network (Newman 2018).

To a large extent, a swarm is a mobile network with flexible architectures—as in a conventional network—but with a robot and sensors at the nodes and AI controlling the functionality and operations of the swarm itself and its objectives.

Issues with Mini-Drones

The proliferation of drones has raised a vast number of issues with their use, efficacy, ethical implications, control, and specific impacts on conflict. The following sections examine five of these issues.

Issue 1: The implications of mini-, micro- and nano-drone swarms

UAVs come in various sizes and have the ability to fly and hover, and range from the cheap Dragon to more expensive Black Hornet. Many of these are designed around insect structures and can capture video from over a kilometer away. The lower-cost drones can be used in a swarm and, with the software constructed, combined with the sensor inputs could emulate the performance of the more expensive drones. The RoboBee is only three centimeters across and weighs 80 milligrams (Grossman 2018). These ultra-small drones are useful for covert surveillance as single entities (often over short distances) and can be flown inside buildings. In swarms, they can have the same effect but are more likely to be spotted. Whilst the engineering for these devices is well advanced, the real secret is the AI driving the swarm coordination and sensors. Hambling (2015) puts a large emphasis on the development of the software, using such techniques as neural networks to improve their performance. These swarms would be harder to spot and could deliver not only good intelligence, but targeted physical effects as well. Of course, these swarms can be used to inspect hazardous environments, such as nuclear accidents, bush fires, and marine environments, by, for example, carrying a line between vessels in stormy conditions. The larger ones can be used as a cheaper and more specific option to deliver materials, such as food, where people are isolated because of fires, avalanches, and epidemics. The micro- and nano-drones can also be used for inspection of injured victims.

Issue 2: Other types of drone swarms and their implications

Sanders (2017) classifies drone swarms into three categories:

- Aerial
- Ground
- Maritime

These categories can be further subdivided by Maritime (Surface), Maritime (Subsurface), and Space. In the aerial environment, Sanders (2017) uses Russia's six-generation fighter jet with five to ten unmanned drones to increase intelligence and targeting. These can leave the atmosphere. The U.S. and NATO have comparable capacity. Employing an example from the Chinese use of drones, Sanders (2017) states that functionality comes from decoying, jamming, and general electronic warfare, as well as kamikaze attacks. In other words, many nations are using them for multiple uses.

Sanders (2017) states that ground drones will be "ubiquitous, self-organized and collaborative". Of course, they can function without any rest. Most of these drones are larger than the mini- drones, discussed above. However, the concept and application of 'smart dust' which consists of wireless networked MicroElectromechanical Systems (MEMs) is comparable to that of a ground-based swarm (Marr 2018). These tiny, grain-sized devices can collect data on such things as moisture and sound, and can store it and/or send it back to a base.

In the maritime environment, various drone swarms have been developed for surface and amphibious craft as well as swarms for submarines that are analogous to those associated with fighter aircraft—that is, to collect intelligence around inshore areas and ports as well as data needed for the safety of the submarine mother ship. The UK has developed a fully functional Unmanned Underwater Vehicle (UUV) for the Royal Navy to act as a substitute for Hunter-Killer manned submarines (Jagger 2020).

Issue 3: Will the development of small drone swarms encourage the advancement changes in conflict tactics such as asymmetric war and terrorism?

As the development of state-based war often leads towards an asymmetric format, the distinctions between both the 'conventional' and civilian conflicts seem to be merging. The Russian use of Hybrid Warfare (Galeotti 2016), the Chinese use of the 'Unrestricted Warfare' doctrine (Liang & Xiangsui 2015) and the American doctrine of the Third Offset (Center for Strategic and International Studies [CSIS] 2017) have led to a situation that avoids major battles or outbreaks of violence. The methods of conflict seem to be moving to be more of an emphasis on 'softer' approaches, such as cyberwar, information warfare, and influence operations. These show similarities to counter-insurgency situations and, since 9/11, the militarization of law enforcement has increased. Korpela (2017) points out that demographic movements will encourage the growth of larger and heavily populated cites. This environment is conducive for the use of mini-drones for real-time surveillance for both low-level civil violence and the distribution of leaflets, and crowd control chemicals. Of course, the relative low cost of these functionally relevant drones encourages all sides to use them. This is aided by the general skills base in flying drones learned from electronic games and hobby drones as well as familiarity with software and hardware of these robots. In an asymmetric warfare scenario, mini-, micro-, and nano-drones have significant flexibility to provide surveillance, psychological, and kinetic functionality.

Issue 4: How will micro- and nano-drone swarms increase social surveillance?

The amount of surveillance of the public has increased significantly in recent times in various countries. As mentioned above, drones can be used by almost anyone to obtain legal and illegal data. The use of mini- and micro- drones by Chinese authorities for their Social Credit System (Kobie 2019) is well recorded. Although in the West the surveillance is not so oppressive, it is increasing in its coverage. Phenomena such as Smart Cities, traffic monitoring, and CCTV coverage provide, in some instances, an oppressive public environment. The Internet of Things (IoT) (Bhuvaneswari & Porkodi 2014) has certainly provided an interface between physical devices and the Internet. This provides a platform for surveillance. The coupling of drone swarms into the IoT has provided some detrimental effects as, in a sense, it is providing a conflict situation even when the stated aim is seemingly legitimate, such as monitoring traffic, high streets, or troublesome housing estates. In overt areas of conflict, swarming drones are known to have detrimental effects on its victims (as well as benefits for the owners of the system). Nemar (2017) describes the potential psychological distress caused by the constant surveillance as well as the potential for harm done by swarms of any size given the kinetic impact they can cause. While this might seem a far cry from some 'innocent' civilian networks of face-recognition cameras in local streets and network-linked devices that control such things as water flow in a reservoir, pouring molten metal in a smelter, a satellite-driven navigation device in a truck, or a remote-control initiator of a domestic heater, the potential to create havoc is still there. This is especially true as the IoT is not renowned for its security prowess (Etter 2016).

Issue 5: Will autonomy of drone swarms especially (LAWs) cause ethical dissonance?

This issue is debated continuously among those who see it as a development in a competitive arms race between various nations and other groups (for instance, Brock 2017) and those who prefer to challenge this on ethical grounds. Although a drone's capability is determined by its software/algorithms, its suite of sensors, and generally its automatic systems are used to control the machine by instructions from the AI driver. An autonomous drone would have the 'human-out-of-the-loop' scenario, where the drone itself would make independent choices about the target-and-kill command. Normally, a human pilot would make the choice; or, if it was automatic, then the (supposedly) built-in instructions would limit the drone to the conditions for which it was designed. However, Artificial General Intelligence (AGI) is different because it assesses the situation from its sensors and can change the algorithms that control it by machine learning. Therefore, over time, the initial designers of the drones might not know what drives the conditions in which lethal force will be used. This is a dilemma as autonomous drones might be needed for very rapid responses. Is it an ethical activity in war to allow the decision of life and death to be made at the discretion of a machine?

A 2019 report to the U.S. Congress (Congressional Research Service 2019) has implied that AI will be used in many aspects of conflict: Intelligence, Surveillance, Reconnaissance, Information Operations, Cyber Operations, and Command and Control. There are some who are extremely optimistic about the future of these aspects, for example, Hambling (2015) who thinks the development of AI for drone swarms will increase exponentially as Moore's Law did with microprocessors. Others like Cross (2020) are a bit more reticent about the future abilities of drones to be used in conflict. Whatever the future of drones, Payne (2018) feels that they will be a significant factor in the future of conflict. Will the inner conflict between accidental killing and the efficiency of the drone swarm as a weapon cause any dissonance?

Conclusion

The concept of a swarm inherently drives drone development toward autonomy. Smart drone swarm technology could have a significant impact on every area of military capability, from enhancing supply chains to Command, Control, Computers, Communications, Cyber, Intelligence, Surveillance and Reconnaissance (C5ISR) and delivering kinetic ammunitions. Swarms of small attack drones that confuse and overwhelm antiaircraft defence could soon become an important part of the modern military arsenal; and, as Britain's defence secretary has said, it is something that would "mark a major evolution in robot-enabled warfare" (Gregg 2019a).

The fact that the components of the swarm can communicate with one another makes the swarm different from a group of individual drones. Smart communication and autonomy allow the swarm to adjust behavior in response to real-time information. Drones equipped with cameras and other environmental sensors (sensor drones) can identify potential targets, environmental hazards, or defenses and can relay that information to the rest of the swarm. The swarm may then maneuver to avoid a hazard or defense, or a weapon-equipped drone (attack drone) may strike the target or defense. Real-time information collection makes drone swarms well-suited for searching over broad areas for mobile or other hard-to-find units in military or civilian operations.

While individual drones can be useful, a swarm of them would be more difficult to eliminate. A swarm of drones would help with a complicated environment, like an urban or covered terrain, where it is hard to see long distances. A large group of drones can provide better situational awareness than single drone.

According to Kallenborn (2018) a future drone swarm need not consist of the same type and size of drones but could incorporate both large and small drones equipped with different payloads. Joining a diverse set of drones creates a whole that is more capable than the individual parts. A single drone swarm could even operate across domains, with undersea and surface drones or ground and aerial drones coordinating their actions.

Swarming also adds new vulnerabilities. Drone swarms are particularly vulnerable to electronic warfare attacks. Because drone swarms are dependent on drone-to-drone communication, disrupting that signal also disrupts the swarm. As swarms become more sophisticated, they will also be more vulnerable to cyberattack. Adversaries may attempt to hijack the swarm by, for example, feeding it false information, hacking it, or generating manipulative environmental signals (Kallenborn & Bleek 2019)

How can an entity defend against a drone swarming attack? The US Air Force unveiled a new tool for that: a high-powered microwave system called Tactical High-Power Microwave Operational Responder (THOR), which is designed to protect bases against swarms of drones. According to the USAF, this system is designed to take out a large number of drones all at once and has a further range than bullets or nets (Cohen 2019). The counterattacks needed against small and very small drone swarms will need to be manifest as there are so many types and volumes of single drones present, as well as a multitude of configurations, sensors, physical environment of conflict, and swarm performance abilities to make it difficult to counter their missions. Of course, many of these swarms have much in common but they also have multiple structures and operational environments and aims and the ability to change their structures, sensors, and AI functions. It is unlikely to be an easy task to counter a well-structured swarm.

The variety of small drones combined with swarm configurations is enormous and, with such a flexible tool, applications are likely to be very diverse and to have an enormous impact on the practices used in conflict.

It is significant that on the first page on Wittes and Blum’s (2016) text on the Future of Violence they paint a scenario of Do-It-Yourself (DYI) built drones to spread deadly spores over the country. Although drone swarms are not mentioned, the use of a swarm could be disastrous.

Drone Swarms are sophisticated combinations of hardware and software. They are almost universal in their potential benefits; however, this is true for their potential dangers as well.

References

Altigator Unmanned Systems 2019, viewed 3 Oct 2019, <https://altigator.com/drone-uav-uas-rpa-or-rpas/>.

Arquilla, J & Ronfeldt, D 2000, *Swarming and the future of conflict*, RAND, Santa Monica, CA, US.

Bhuvaneswari, V & Porkodi, R 2014, ‘The Internet of Things (IoT) applications and communication enabling technology standards: An overview, *Proceedings of the 2014 International Conference on Intelligent Computing Applications*, Coimbatore, Tamil Nadu, IN, viewed 20 June 2019, pp. 324-9, <http://ieeexplore.ieee.org/abstract/document/6965065>, doi: 10.1109/ICICA.2014.73.

Brock, JW 2017, *Why the United States must adopt lethal autonomous weapons systems*, United States Army Command and General Staff College, Fort Leavenworth, KS, US.

Chung T 2016, ‘OFFensive Swarm-Enabled Tactics (OFFSET)’, viewed 5 October 2019, <https://www.darpa.mil/program/offensive-swarm-enabled-tactics>.

Cohen RS 2019, 'Microwave weapons moving toward operational use', *Air Force Magazine*, 20 March, viewed 15 October 2019, <https://www.airforcemag.com/microwave-weapons-moving-toward-operational-use/>.

Congressional Research Service 2019, *Artificial Intelligence and national security*, updated 21 November 2019, Washington DC, US.

Cross, T 2020, 'Reality check', *The Economist Technical Quarterly*, 13 June, pp.3-4.

Center for Strategic and International Studies (CSIS) 2017, *Assessing the Third Offset Strategy*, Center for Strategic and International Studies, March, Washington, DC, US.

Department of Defense (DoD) 2005, *Unmanned Aircraft Systems roadmap 2005-2030*, Office of the Secretary of Defense, 20 July, Pentagon, Arlington, VA, US.

——2014, *Unmanned Systems Integrated Roadmap 2013-2038,* Under Secretary of Defense Acquisition, Technology & Logistics, January, Pentagon, Arlington, VA, US.

——2018, *Unmanned Systems Integrated Roadmap 2017-2042*, Office of the Secretary of Defense, 28 August, Pentagon, Arlington, VA, US.

Etter, D 2016, *IoT Security: Practical guide book*, Lightning Source, Milton Keynes, UK.
Finland Ministry of Defence (MoD) 2015, *Preliminary assessment for replacing the capabilities of the Hornet Fleet final report*, Ministry of Defence, 08 June, Helsinki, FI.

Galeotti, M 2016, *Hybrid War or Gibridnaya Voina? Getting Russia's non-linear military challenge right*, Mayak Intelligence, Milton Keynes, UK.

George, P 2019, 'Artificial intelligence system 'too good' to be released but drone development continues', Your NZ, 20 February, viewed 21 September 2020, <https://yournz.org/2019/02/20/artificial-intelligence-system-too-good-to-be-released-but-drone-development-continues/>.

Gregg A 2019a, 'A key U.S. ally is close to adding swarming attack drones to its military arsenal', *Washington Post*, 15 February, viewed 21 September 2020, <https://www.washingtonpost.com/business/2019/02/15/key-us-ally-is-close-adding-swarming-attack-drones-its-military-arsenal/>.

——2019b, 'Swarming attack drones could soon be real weapons for the military', *Washington Post*, 19 February, viewed 5 October 2019, <https://www.latimes.com/business/la-fi-drone-swarms-20190219-story.html>.

Grossman, N 2018, *Drones and terrorism: Asymmetric warfare and the threat to global security*, Taurus and Co., London, UK.

Haberl F & Huemer F 2019, 'The terrorist/*jihadi* use of 3D-printing technologies: Operational realities, technical capabilities, intentions and the risk of psychological operations', *Proceedings of the 14th Annual International Conference on Cyber Warfare and Security* (*ICCWS 2019*), 28 February-1 March 2019, Stellenbosch, ZA.

Hambling, D 2015 *Swarm troopers*, Archangel Ink. Venice, FL, US.

International Civil Aviation Organization (ICAO) 2019, *Remotely Piloted Aircraft System (RPAS) Concept of Operations (CONOPS) for international IFR operations*, viewed 11 October 2019, <https://www.icao.int/safety/UA/Documents/ICAO%20RPAS%20CONOPS.pdf>).

Jagger,S 2020, 'Royal Navy to field large "robot subs"', *Warships International Fleet Review*, p 34.

Joint Air Power Competence Centre (JAPCC) 2014, 'Remotely Piloted Aircraft Systems in contested environments: A vulnerability analysis', September, viewed 12 October 2019, <https://www.japcc.org/portfolio/remotely-piloted-aircraft-systems-in-contested-environments-a-vulnerability-analysis/>.

Kallenborn Z 2018, *The era of the drone swarm is coming, and we need to be ready for it*, Modern War Institute at West Point, 25 October, viewed 17 October 2019, <https://mwi.usma.edu/era-drone-swarm-coming-need-ready/>.

——& Bleek PC 2019, 'Drones of mass destruction: Drone swarms and the future of nuclear, chemical, and biological weapons', *The Nonproliferation Review*, vol. 25, nos. 5-6, Special section on the nuclear dimensions of the 1967 Arab–Israeli War, pp. 523-43, 2 January, viewed 2 January 2019, <https://doi.org/10.1080/10736700.2018.1546902>.

Kobie, N 2019, 'The complicated truth about China's social credit system', *Wired*, viewed 8 July 2020, <https://www.wired.co.uk/article/china-social-credit-system-explained>.

Korpela, C 2017, 'Swarms in the Third Offset', ed. White, S.R, *Closer than you think: The implications of the Third Offset Strategy for the U.S. Army*, US Army Command and General Staff College, Fort Leavenworth, KS, US.
Liang, Q & Xiangsui, W 2015, *Unrestricted warfare: China's master plan to destroy America,* Echo Point Books & Media, New York, NY, US.

Marr, B 2018, 'Smart dust is coming. Are you ready?', *Forbes,* viewed 8 July 2020, <https://www.forbes.com/sites/bernardmarr/2018/09/16/smart-dust-is-coming-are-you-ready/#4b099ec05e41>.

McMullan T 2019, 'How swarming drones will change warfare', BBC News, 16 March, viewed 15 October 2019, <https://www.bbc.com/news/technology-47555588>.

Nemar, R 2017, 'Psychological harm', *The Humanitarian impact of drones*, eds. R Acheson, W Bolton, E Minor & A Pytlak, Women's International League for Peace and Freedom, International Disarmament Institute, New York, NY, US, pp. 36-47.

Newman, M 2018, *Networks,* 2nd edn., Oxford University Press, Oxford, UK.

Papireddy, T 2015, *Tracking and monitoring Unmanned Aircraft Systems activities with crowd-based mobile apps*, 1 May, School of Computer Science, Howard R. Hughes College of Engineering, University of Nevada, Las Vegas, NV, US.

Payne, T 2017, ‘Lethal autonomy: What it tells us about modern warfare’, *Air & Space Power Journal*, vol. 15, no. 14, Winter, pp. 16-33.

——2018, *Strategy, Evolution, and War*, Georgetown University Press, Washington, DC, US.

Peters, J 2019, ‘Watch DARPA test out a swarm of drones’, *The Verge*, 9 August, viewed 18 October 2019, <https://www.theverge.com/2019/8/9/20799148/darpa-drones-robots-swarm-military-test>.

Pritchard S 2019, ‘Drones are quickly becoming a cybersecurity nightmare’, *Threatpost,* vol. 22, March, viewed 15 October, <https://threatpost.com/drones-breach-cyberdefenses/143075/>.

Pennsylvania State University 2019, ‘Classification of the Unmanned Aerial Systems’, University Park, PA, US, viewed 12 October 2019, <https://www.e-education.psu.edu/geog892/node/5>.

PricewaterhouseCoopers (PwC) 2016, ‘Global market for commercial applications of drone technology valued at over $127bn’, *PwC blog*, 9 May, viewed 11 October 2019, <https://pwc.blogs.com/press_room/2016/05/global-market-for-commercial-applications-of-drone-technology-valued-at-over-127bn.html>.

Sanders, AW 2017, *Drone swarms*, School of Advanced Military Studies, Fort Leavenworth, KS, US.

Singer PW 2009, *Wired for war: The robotics revolution and conflict in the 21st century*. Penguin Books, London, UK.

Single European Sky ATM Research (SESAR) 2016, European Drones - Outlook Study -Unlocking the value for Europe, 1 November, European Union and Eurocontrol, Brussels, BE.

Wittes, B & Blum, G 2016, *The future of violence: Robots and germs, hackers and drones*, Amberley, Stroud, UK.

Human Rights and Artificial Intelligence: A Universal Challenge

VA Greiman

Boston University
Boston, Massachusetts, United States

E-mail: ggreiman@bu.edu

Abstract: *As artificially intelligent systems benefit citizens around the globe, there remain many ethical questions about the intrusion of AI into every aspect of our private and professional lives. This paper raises awareness of the unprecedented challenge that governments and private industry face in managing these complex systems that include regulators, markets, and special interests. The research focuses on three primary areas: (1) how to determine whether or when various applications of AI are ethical, or whether they violate basic human rights; (2) who should make these decisions concerning the use of AI in the public and business environment, and who should be held accountable when things go wrong; and (3) how to improve the accountability frameworks and regulations essential to ensure safety and security in advancing artificially intelligent systems.*

Keywords: *Artificial Intelligence, Human Rights, Ethics, Machine Learning, Algorithms*

Introduction

In recent years, there has been a recognition that the increasing presence of Artificial Intelligence creates enormous challenges for human rights and also gives a new relevance to moral debates that used to strike many as arcane (Risse 2019). The potential benefits of AI as demonstrated in the research are tremendous. AI could play a crucial role in many aspects of human problems, including climate change, international conflict, and medical breakthroughs, and could contribute greatly to improving human exploration, scientific knowledge, and even quality of life. However, the research on AI and human rights has raised a number of issues where AI deployment has created serious concerns about its impact on decisions and outcomes (Raso *et al.* 2018; Yu 2019). As of 2019, according to the International Data Corporation (IDC), worldwide spending on AI and cognitive systems is set to triple in the near future. In fact, investment is predicted to grow from 37.5 billion U.S. dollars in 2019 to 97.9 billion U.S. dollars by 2023 (Seeley 2019). Moreover, Price Waterhouse Coopers (PwC) projects that Artificial Intelligence will contribute 15.7 trillion U.S. dollars to the global economy by 2030 (PwC 2017). In the United States, according to budget documents released March 18, 2019, the federal government is preparing to invest about $4.9 billion in unclassified Artificial Intelligence and machine-learning-related research and development in fiscal 2020 (Cornillie 2019). The White House R&D budget request spans civilian agencies led by the Energy Department, NIH, NIST, and NSF, along with defense spending needs. The Department of Defense's research will focus on areas such as using AI and machine-learning techniques as part of cybersecurity solutions to safeguard the nation's power grid or developing next-generation microelectronics for applications in AI-enabled devices (DoD 2018).

ISSN 1445-3312 Print/ISSN 1445-3347 Online

More than a dozen countries have launched AI strategies in recent years, including China (US-China 2019), France (2017), Canada (2018), and South Korea (Republic of South Korea 2017). Their plans include items such as new research programs, AI-enhanced public services, and smarter weaponry. On February 11, 2019, President Trump signed Executive Order 13859 announcing the American AI Initiative—the United States' national strategy on Artificial Intelligence. This strategy is a concerted effort to promote and to protect national AI technology and innovation. The Initiative implements a whole-of-government strategy in collaboration and engagement with the private sector, academia, the public, and like-minded international partners. It directs the federal government to pursue five pillars for advancing AI: (1) promote sustained AI R&D investment, (2) unleash federal AI resources, (3) remove barriers to AI innovation, (4) empower the American worker with AI-focused education and training opportunities, and (5) promote an international environment that is supportive of American AI innovation and its responsible use (White House 2019).

Investment in AI cuts across industries and has been particularly prevalent in the large technology-driven companies including IBM, Google, Apple, Intel, Amazon, Microsoft, and Uber. IBM's most notable investment in AI is Watson, a cognitive computing platform that can answer questions posed in natural language to extract meaning from photos, videos, text, and speech. In September 2016, Microsoft created an Artificial Intelligence and Research Group that cuts across the Windows, Office, and Azure business units. In 2019, Microsoft revealed plans to invest $1 billion in OpenAI, a San Francisco-based company, and also announced a two-year partnership to develop AI supercomputing technologies on Microsoft Azure (Langston 2020; Techworld 2019).

Since the first design of a programmable machine in 1837 by Charles Babbage and Ada Lovelace, AI has been transforming the economies of every nation on the globe in ways that create challenges and opportunities for the advancement of AI insofar as it impacts jobs, improves healthcare, provides autonomous transportation vehicles, and enhances national security. Potentially, the most important question raised by this transformation is how will AI interact with international security (Congressional Research Service 2019)? Other questions also arise: What are the near-term security challenges (and opportunities) posed by AI? Could AI radically shift key strategic parameters, for instance, by enabling powerful new capabilities (in cyber, lethal autonomous weapons, military intelligence, strategy, science), by shifting the offense-defense balance, or by making crisis dynamics unstable, unpredictable, or more rapid (Dafoe 2018)? What is the impact that recent advances in AI could have on strategic stability and nuclear risk (Boulanin 2019)? Could trends in AI facilitate new forms of international cooperation, such as by enabling strategic advisors, mediators, or surveillance architectures or by massively increasing the gains from cooperation and costs of non-cooperation? If general AI comes to be seen as a critical military (or economic) asset, under what circumstances is the state likely to control, close, or securitize AI R&D? Also, what are the conditions that could spark and fuel an international AI race (Dafoe 2018)?

Beyond technical aspects, the emergence of AI has raised many new legal, policy, and regulatory challenges, broad and complex in scope. This paper will provide a deeper understanding of AI technologies and the risks that AI, algorithms, and machine learning may pose to human rights. The question of how AI can be regulated to provide redress for individuals and groups that are adversely affected by algorithmically informed decisions will be addressed. Defining the effect of AI technologies on developing countries including inequality, breach of privacy, and discriminatory treatment is central to this research, along with identifying responses and approaches to regulating AI. The paper will also highlight the legal and ethical implications of this new technology on human rights and economic development. Finally, the paper will also stimulate ideas for managing the challenges and risks in AI deployment.

The Development of Artificially Intelligent Systems

In a statement before the United States (US) Senate Commerce Subcommittee on Space, Science, and Competitiveness, Technical Fellow and Director of Microsoft Research - Redmond Lab, Eric Horvitz explained that "Artificial intelligence (AI) refers to a set of computer science disciplines aimed at the scientific understanding of the mechanisms underlying thought and intelligent behavior and the embodiment of these principles in machines that can deliver value to people and society" (2016). Accountability for decisions of government and private industry that involve such important rights as due process, the rule of law, eligibility for welfare and social services, and employment permeate every aspect of our lives. Now and in the future, many of these decisions will be determined by data and algorithmic decision making. Technically, AI proceeds through phases of development determined by the perspectives of the designers. **Figure 1**, below, shows the development of AI from initial data collection and analysis to algorithmic modelling through machine learning. The following sections of this article look at the process of data collection and analysis (which is the foundation of all AI) and at its impact on the following phases of AI development.

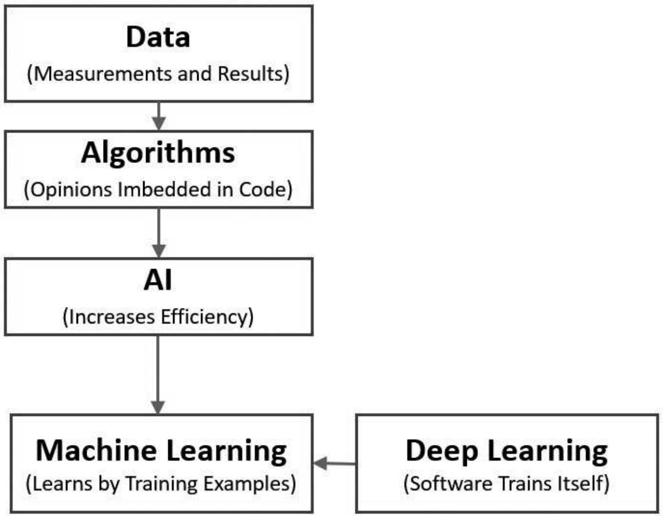


Figure 1: Artificial Intelligence development phases

Data Collection

One of the major challenges for the implementation of AI is the collection and the analysis of vast amounts of data. AI can only be as good as the data it is given. If the input data reflects biases, so too will the output (Kleinberg *et al.* 2017). The regularity of the data is dependent on a variety of factors, including the variance in published information, the completion of the published information, the consistency of the data collected, and any bias that might exist in the development of the data. An interdisciplinary approach is essential to understanding the various impacts that data can have on human rights, privacy, and confidentiality (Raso *et al.* 2018). To the extent that the data used to 'train' an AI system is biased, the resulting system will reflect, or perhaps even exacerbate, those biases (Osoba & Welser 2017). This can have profound consequences for a wide variety of human rights—depending on what the system is intended to do. Decisions made by an AI system's human designers can have significant human rights consequences. Human designers can, for example, prioritize the variables they would like the AI system to optimize and decide what variables the AI should take into consideration as it operates. Such design decisions can have both positive and negative human rights impacts, which will be informed by the individual life experiences and biases of the designers (Raso *et al.* 2018). Expert analysis and input from various disciplines should be employed to provide a broader perspective to the data analysis. The interaction between the AI analysis and human analysis is critical to the development of predictive accuracy that can better support rational decision making (Agrawal, Gans & Goldfarb 2017). Some important questions in data analysis include (1) What are the sources of the data and will they be sufficient to address the problem identified as suitable for an AI analysis? (2) Who will collect and compile the data and ensure its legitimacy and authenticity? (3) How will the quality of the data be tested before building the machine-learning algorithms?

Algorithmic Modelling and Machine Learning

The next step is the analysis of the data and the modelling of the data. This requires that the data be analysed both on a qualitative and quantitative basis using algorithmic modelling (Kleinberg *et al.* 2017). Algorithmic models have been used for decades in the insurance industry to assess aggregate exposure and can be used as models for other disciplines. Algorithms are the building blocks that make up machine learning and Artificial Intelligence, whereas machine learning (sometimes called deep learning) is a subset of AI composed of algorithms that allows computers to learn without being explicitly programmed (Linux 2005). Algorithms can impact people's lives without anyone knowing that anything has happened. In addition to knowing people's search results and all kinds of personal preferences—such as travel, entertainment, and religion—algorithms can predict people's health (Marr 2018), follow people's everyday activities (Stanford University 2016); increase inequality and threaten people's democracy (O'Neil 2016); and determine if people have criminal charges or sentences (Angwin *et al.* 2016; Wexler 2017). The effectiveness of algorithms is increasingly enhanced through 'Big Data': the availability of an enormous amount of data on all human activity and other processes in the world. Such data allows a type of AI known as 'machine learning' to draw inferences about what happens next by detecting patterns. Any system designed by humans reflects human bias (Risse 2019). Big data may eventually outdo humans on everything. Decisions such as what college to attend, where to live, when to travel and where to go, how to stay healthy, what employment opportunities exist, and what job to take, or who qualifies for these jobs may eventually be a decision for the machine rather than the human.

Machine learning is a subfield of Artificial Intelligence and computer science that allows software applications to be more accurate in predicting results. Deep learning is a subfield of machine learning and is what powers the most human-like Artificial Intelligence (Goodfellow, Bengio & Courville 2017). Deep learning also uses training data to find patterns used to make predictions about new data, though in this case features are not extracted by humans. Rather, data sets are fed directly into the deep learning algorithm, which then predicts the occurrence of objects. Deep learning algorithms do this via multiple layers of artificial neural networks that mimic biological brains (Livingston & Risse 2019). Output data of one layer is input data for the next layer, and so forth. Deep learning architecture is used in speech recognition, natural language processing, and computer vision (Chui *et al.* 2018; Goodfellow, Bengio & Courville 2017; Kelly 2014).

Moreover, Artificial Intelligence can be categorized into three basic stages of development. Basic AI or Artificial Narrow Intelligence (ANI) is limited in scope and restricted to just one functional area. AlphaGo, a computer program that plays the board game Go, is an example. Advanced AI or Artificial General Intelligence (AGI) usually covers more than one field, such as power of reasoning, abstract thinking, or problem solving on par with human adults. Autonomous AI or Artificial Super Intelligence (ASI) is the final stage of intelligence expansion in which AI surpasses human intelligence across all fields. This stage of AI is not expected to be fully developed for several decades (Mou 2019).

Human Rights and International Law

Although interpreted in different ways, the human rights framework rests on decades of global consensus and has been imbedded in constitutions and applied by governments around the world in national regulations; this framework also establishes a universally applicable set of norms and commitments (Donahoe & Metzger 2019).

A broad definition of human rights on the global scale can be found in the Universal Declaration of Human Rights (UDHR) adopted as a non-binding resolution. However, Canada and many other sovereign states have long believed there is an obligation for states to observe the human rights and fundamental freedoms set forth in the UNDHR (1948) that derives from their adherence to the Charter of the United Nations. States are not alone in observing these rights, as businesses have also accepted their own responsibilities to protect human rights. These responsibilities have been described in the United Nations Guiding Principles on Business and Human Rights (United Nations 2011) also known as the Ruggie Principles because of their drafter, Harvard Professor John Ruggie. The Ruggie Principles consist of 31 directives, framed in three main pillars: the state duty to protect against human rights abuses; the corporate responsibility to respect human rights; and the need to help victims achieve remedy. Supported by these three pillars, the Ruggie Principles propose that companies, in order to demonstrate and implement respect for human rights, take measures such as (1) making a public commitment to respect human rights; (2) identifying, mitigating, and accounting for damage caused to human rights; and (3) disposing of procedures for remedying the negative consequences on human rights they cause or contribute to causing (United Nations 2011). Although the Ruggie Principles do not have the force of law, they clarify how pre-existing international rights' standards apply to business activities and provide guidance to businesses on how they can operate in a rights-respecting manner (Raso *et al.* 2018). Giving effect to these expectations raised in the Ruggie Principles requires the establishment of monitoring and oversight mechanisms that apply throughout the entire algorithmic process (McGregor, Murray & Ng 2019).

On a transnational scale, the European Union has been proactive in its approach to AI and the development of governance frameworks to address the ethical concerns in the use of AI. The European Parliament's 2017 report on Civil Law Rules on Robotics is one example. While this resolution is not binding, it does express the Parliament's opinion and makes various requests of the European Commission to carry out further work on the topic. In particular, the Resolution "consider[ed] that the existing Union legal framework should be updated and complemented, where appropriate, by guiding ethical principles in line with the complexity of robotics and its many social, medical and bioethical implications" (European Parliament 2017). Set out in its Annex are a proposed Code of Ethical Conduct for Robotics Engineers, Code for Research Ethics Committees, License for Designers, and License for Users. The Parliament also requested the European Commission to submit a "proposal for a legislative instrument on legal questions related to the development and use of robotics and AI foreseeable in the next 10 to 15 years, combined with non-legislative instruments such as guidelines and codes of conduct as referred to in recommendations set out in the Annex" (European Parliament 2017). As this remains a work in progress, the outcome may not be known for several years.

At the international level, the United States is one of 42 Countries to have adopted the Organization for Economic Cooperation and Development (OECD) AI Principles. The principles promote Artificial Intelligence (AI) that is innovative and trustworthy and that respects human rights and democratic values (OECD 2019). In June 2019, the G20 adopted human-centred AI Principles that draw from the OECD AI Principles (G20 2019). Because of the wide acceptance of these principles they have become an important foundation for the development of legal and regulatory frameworks.

Improvement of our Frameworks for AI Regulation

If governments are going to rely upon AI, they must develop processes to evaluate how machine results affect people's rights and must be ready to provide timely, effective remedies for cases in which machine-made decisions turn out to have been wrong (Donahoe & Metzger 2019). Initiatives by Google (2020) and Microsoft to develop their own ethical principles for AI and collaborative efforts between private companies and other stakeholders including the Partnership on AI, the Asilomar Principles (Future of Life Institute [FLI] 2017), and OpenAI are a good start, but more collaborative efforts are needed between industry and government to identify policy and regulatory frameworks to prevent harmful outcomes in machine learning (World Economic Forum [WEF] 2018).

The regulation of AI is a topic not extensively covered in the law, public policy, or ethical journals, and further development of the understanding of the impacts of AI on human rights is desperately needed. The traditional methods of regulation must be reviewed and analysed to determine how AI is different and what approaches are needed to regulate AI, as well as who should be held accountable when rights are violated. Though beyond the scope of this paper, this requires an analysis of legislative approaches, international resolutions and treaties, self-regulation and standardization, certification systems, contractual rules, soft law, and agile governance among other regulatory approaches (Walz & Firth-Butterfield 2019).

In May 2018, Amnesty International, Access Now, along with several partner organizations, launched the Toronto Declaration on protecting the right to equality and non-discrimination in Machine Learning Systems (MLS) (Open Democracy 2018). The Toronto Declaration has three primary goals easily adaptable by other government organizations. The first sets out the duties of states to prevent discrimination in the context of designing machine learning systems in public contexts or through public-private partnerships. This includes identification of risks, including mechanisms for independent oversight. The second outlines the responsibilities of private actors in the context of the development and deployment of ML systems and considering the use of third- party independent auditors when there is a significant risk of human rights abuses. The third Declaration asserts the right to effective remedy and accountability by the abuser. This requires governments to outline clear lines of accountability and clarification as to which government bodies or individuals are legally responsible for decisions made through the use of such systems.

Various approaches to AI regulation have been offered by governments and private industry, and continued dialogue and collaboration are essential. For example, the U.S. Public Policy Council, in its recommendations for governance of AI, has proposed adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically-informed decisions along with accountability of institutions for decisions made by algorithms they use, even if it is not feasible to explain in detail how the algorithms produce their results (Association for Computing Machinery [ACM] 2017). Another important approach has been recommended by IEEE's Global Initiative on the Ethics of and Autonomous and Intelligent Systems, which calls for an ethics- and values-based approach to dealing with the impacts of intelligent and autonomous systems that prioritize human well-being within a given cultural contest (IEEE 2017). These initiatives and others like it can play a critical role in framework development to address current gaps in the laws and regulatory process.

Given the extent of the impact of AI on society, it is unsurprising that the use of algorithms in decision making raises a number of human rights' concerns, creating a call for a rigorous accountability framework (McGregor, Murray & Ng 2019). A key consideration in the use of algorithms is not only whether safeguards are in place, but also whether they are able to operate effectively (McGregor, Murray & Ng 2019). A critical court decision that involved the use of algorithmic risk assessments in sentencing decisions was raised by the Wisconsin Supreme Court in the case of the State of Wisconsin *v.* Eric L. Loomis (Wisconsin *v.* Loomis 2016). In this case, the court had to assess whether the use of an algorithmic risk assessment tool to determine if the defendant could be supervised within the community rather than detained violated the defendant's right to due process. As the court noted, risk scores are intended to predict the general likelihood that those with a similar history of offending are either less likely or more likely to commit another crime following release from custody. However, the court found that the risk assessment used in this case did not predict the specific likelihood that an individual offender will reoffend. Instead, it provides a prediction based on a comparison of information about the individual to a similar data group (State of Wisconsin *v.* Loomis, 2016). The following section focuses on how algorithms impact human rights and who should be held accountable for the violations they cause.

Who Should Be Accountable When AI Decision Making Goes Wrong?

The question of who should be accountable when AI decision making goes wrong has been raised by scholars as a major concern in the use of AI. Robots cannot be equated with human innovators and creators. Their 'byproducts' are not generated in a way that resembles human creativity. Thus, IP laws must not allow for granting IP rights to robots, and their byproducts should forever rest in the public domain (Khoury 2017). This paper proposes that accountability for algorithmic decision making should be linked not just to existing human rights laws, principles, and frameworks, but also to new initiatives by governments and that the private sector should also be advanced to address the growing gap between AI development and the law. Algorithmic accountability means defining in the law, contracts, and oversight frameworks the responsible parties and then providing an identity tag for each AI system in order to maintain a clear line of accountability (IEEE 2017). Many countries have developed policy and advisory committees to investigate and report to their governments the potential of Artificial Intelligence (Canada 2018; European Union Agency for Cybersecurity [ENISA] 2019; Republic of South Korea 2017). Governments should take seriously the development of accountability frameworks while AI is still in its infancy. Importantly, there may be circumstances in which AI systems should be licensed or certified or tested by an impartial review board and should receive approval before being implemented. For example, an independent review may be necessary where human rights or privacy are at greatest risk, such as jail paroles or criminal sentencing, or recruitment of people for safety-critical or national security positions.

The recently proposed US Algorithmic and Accountability Act in AI is a good example of how policy makers can address the problems of bias and accountability in algorithms (US House 2019). The proposed law will require companies to study and to fix flawed computer algorithms that result in inaccurate, unfair, biased, or discriminatory decisions impacting Americans to take corrective action in a timely manner if such issues are identified. In addition, the bill would require those companies to audit all processes beyond machine learning involving sensitive data—including personally identifiable, biometric, and genetic information—for privacy and security risks. The bill would place regulatory power in the hands of the US Federal Trade Commission, the agency in charge of consumer protections and antitrust regulation. Though the bill is not without its critics, it would allocate responsibility to the parties best able to control the risk of unethical and biased decision making and mitigate or avoid the potential harm to unwary users and consumers (United States House 2019).

AI Driven Call to Action

There has been a call to action both within the United States and the European Union for a more comprehensive, robust approach to AI-driven social impacts. A holistic approach that requires more integration of policies, strategies, and standards at the highest levels of government; more information sharing, collaboration, and mutual assistance at the international level; and development of stronger public-private partnerships among government, the military, industry, and the cyber community.

The United States is one example of a high priority approach that ultimately resulted in the implementation of the federal government’s National Artificial Intelligence Research and Development Strategic Plan, which provides for addressing ethical, legal, and societal considerations in AI, and improving fairness, transparency, and accountability of AI systems by design (WH 2019). Other countries discussed in this paper have enacted policies and strategies that should also be considered. For example, Canada produced a white paper which highlighted the need for an oversight body to review automated decision making, and to provide advice to ministers during the design of AI systems (Canada 2018). The European Commission (EC)n its white paper recommended a coordinated European approach on the human and ethical implications of AI as well as committing to preserving the EU’s technological leadership and to ensuring that new technologies are at the service of all Europeans – improving their lives while respecting their rights (EC 2020). In regard to national defence and AI, a multi-stakeholder-influenced governance or advisory board would afford the opportunity to more smoothly make fundamental turns in the future in defining its goals and priorities. This organization could, thus, avoid the self-sealing morass of long-term, large public initiatives that fail to address the unique dangers lurking in the AI world involving control over a nation's security that leave borders defenceless, and promote the development of control systems to make sure that preventative measures are being used by the industry at large. As discussed in this paper, existing mechanisms provide a starting point for analysis, but more robust, legally-based mechanisms—including treaties, legislation, and independent oversight authority, as well as more information-sharing supported by a regulatory framework—will all help to contribute to a more transparent and enforceable governance structure for the prevention and mitigation of human rights violations.

Conclusion

Challenges raised by organizations universally require a new way of thinking about the governance of AI and approaches to AI design and development. Just as security is designed into Internet and cyber systems, human rights protections must be designed into AI systems. Since humans must be accountable for their decisions, so must Artificial Intelligence and its owners and operators. This article has focused on the link between Artificial Intelligence and global human rights laws and protections. One of the key concerns expressed in this paper is algorithmic accountability. Though various accountability mechanisms currently exist, they do not go far enough. The law of international human rights presents challenges for all organizations, both public and private, and sets forth a framework for making sure those rights are protected. The conclusions reached so far indicate a need for a better understanding of the existing frameworks now used to monitor Artificial Intelligence while being aware of their potential applications—specifically focusing on the challenges of artificial data collection and selection, bias, virtual threats, and systemic risk relating to the different forms of Artificial Intelligence. These conclusions also raise broader issues on protecting property rights, privacy, competition, and confidentiality. One of the greatest challenges for researchers, policymakers, government, and private industry is how to create AI systems that are less artificial, bias free, more intelligent, and humane. If the global development of AI systems is to advance, much work remains to be done to develop accountable frameworks and the legal allocation of responsibility to those who design and operate these intelligent systems.

References

Agrawal, A, Gans, JS & Goldfarb, A 2017, 'What to expect from Artificial Intelligence,' *MIT Sloan Management Review*, vol. 58, no. 3, pp. 22-6.

Angwin, J, Larson, J, Kirchner, L & Mattu, S 2017, 'Minority neighborhoods pay higher car insurance premiums than white areas with the same risk', ProPublica and Consumer Reports, viewed 27 August 2020, <https://www.propublica.org/article/minority-neighborhoods-higher-car-insurance-premiums-white-areas-same-risk>.

Association for Computing Machinery (ACM) 2017, *Statement on algorithmic transparency and accountability*, US Public Policy Council (USACM), 12 January.

Boulanin, V 2019, *The impact of Artificial Intelligence on strategic stability and nuclear risk: Euro-Atlantic Perspectives*, vol. 1, Stockholm International Peace Research Institute, Solna, SE.

Canada Government, Treasury Board of Canada Secretariat 2018, 'Responsible Artificial Intelligence in the government of Canada' Digital disruption white paper series, Version 2.0, 10 April.

Chui, M, Manyika, J, Miremadi, M, Henke, N, Chung, R, Nel, P &Malhotra, S 2018, 'Notes from the AI frontier: Applications and value of deep learning', McKinsey Global Institute, McKinsey and Company, San Francisco, CA, US.

Congressional Research Service (CRS) 2019, 'Artificial Intelligence and national security', R45178, 21 November, U.S. Congress, Congressional Research Digital Collection.

Cornillie, C 2019, 'Finding Artificial Intelligence money in the fiscal 2020 budget', Bloomberg Government, viewed 27 August 2020, <https://about.bgov.com/news/finding-artificial-intelligence-money-fiscal-2020-budget>/.

Dafoe, A 2018, 'AI governance: A research agenda', August 27, Future of Humanity Institute, University of Oxford, UK.

Department of Defense (DoD) 2018, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance our Security and Prosperity*. US Department of Defense, Washington, D.C., US.

Donahoe, E & Metzger, MM 2019, 'Artificial Intelligence and human rights." *Journal of Democracy*, vol. 30, no. 2, pp. 115-26.

European Commission 2020, 'White paper on Artificial Intelligence - A European approach to excellence and trust', 19 February, Brussels, BE.

European Union Agency for Cybersecurity (ENISA) 2019, 'Trustworthy AI requires solid Cybersecurity', 25 October, viewed 27 August 2020, <https://www.enisa.europa.eu/news/enisa-news/trustworthy-ai-requires-solid-cybersecurity/>.

European Parliament 2017, 'European Parliament report with recommendations to the Commission on Civil Law Rules on Robotics', 1 January 2017, Brussels, BE, viewed 27 August 2020, <https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html>.

France 2017, *How can humans keep the upper hand? Report on the ethical matters raised by Algorithms*, 15 December, French Data Protection Authority (CNIL), Government of France.

Future of Life Institute (FLI) 2017, 'Asilomar AI principles', viewed 27 August 2020, <https://futureoflife.org/ai-principles/>.

Goodfellow, I, Bengio, Y & Courville, A 2017, *Deep Learning*, MIT Press, Cambridge, MA, US.
G20 2019, Ministerial Statement on Trade and Digital Economy, 'G20 AI Principles', Annex, G20, Japan, viewed 27 August 2020, http://www.g20.utoronto.ca/2019/2019-g20-trade.html.

Google 2020, 'Artificial intelligence at Google: Our principles,' Google, viewed 27 August 2020, <https://ai.google/>.

Horvitz, E 2016, statement before 'The Dawn of Artificial Intelligence Hearing before the United States (US) Senate Commerce Subcommittee on Space, Science, and Competitiveness', 114th Cong. 2016, 30 November, Washington, D.C., US.

IEEE Global Initiative on Ethics of and Autonomous and Intelligent Systems 2017, 'Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems', viewed 27 August 2020, <https://standards.ieee.org/content/dam/ieeestandards/standards/web/documents/other/ead_v2.pdf>.

Kelly, K 2014, "The three breakthroughs that have finally unleashed AI on the world', *Wired*, viewed 27 August 2020, <https://www.wired.com/2014/10/future-of-artificial-intelligence>.

Khoury, AH 2017, "Intellectual property rights for Hubots: On the legal implications of human-like robots as innovators and creators", *Cardozo Arts and Ent Law Journal,* vol. 35, no. 3, pp. 635-68.

Kleinberg, J, Lakkaraju, H, Leskovec, J, Ludwig, J & Mullainathan, S 2017, 'Human decisions and machine predictions', Working paper 23180, February, National Bureau of Economic Research, NBER Working papers series, Cambridge, MA, US.

Langston, J 2020, 'AI supercomputer: Microsoft announces new supercomputer, lays out vision for future AI work', Microsoft, The AI blog, 19 May, Microsoft, Redmond, Washington, US, viewed 27 August 2020, <https://blogs.microsoft.com/ai/openai-azure-supercomputer/>.

Linux 2005, 'Algorithms: A very brief introduction', 29 July, The Linux Foundation, San Francisco, CA, US.

Livingston, S & Risse, M 2019, 'The future impact of Artificial Intelligence on humans and human rights, *Ethics and International Affairs*, vol. 33, no. 2, pp. 141-58.

Marr, B 2018, 'How is AI used in healthcare?: 5 powerful real-world examples that show the latest advances', *Forbes*, 27 July, viewed 27 August 2020, <https://www.forbes.com/sites/bernardmarr/2018/07/27/how-is-ai-used-in-healthcare-5-powerful-real-world-examples-that-show-the-latest-advances/#111a7cf45dfb>.

McGregor, L, Murray, A & Ng, V 2019, 'International human rights law as a framework for algorithmic accountability', *International and Comparative Law Quarterly*, vol. 68, pp. 309-43.

Microsoft Corporation , viewed 27 August 2020, <https://www.govinfo.gov/content/pkg/CHRG-114shrg24175/html/CHRG-114shrg24175.htm>.

Mou, X 2019, 'Artificial Intelligence: Investment trends and selected industry uses', September, The International Finance Corporation, The World Bank Group, Washington, D.C., US.

O’Neil, C 2016, *Weapons of math destruction: How big data increases inequality and threatens democracy*, Crown Publishing Group, New York, NY, US.

Open Democracy 2018, ‘Human rights and artificial intelligence: The challenge of an era’, Open Democracy, London, UK.

Organization for Economic Cooperation and Development (OECD) 2019, ‘OECD principles on AI’, OECD, Paris, June, viewed 27 August 2020, <https://www.oecd.org/going-digital/ai/principles/>.

Osoba, O & Welser, W 2017, *An intelligence in our image: The risks of bias and errors in Artificial Intelligence*, Rand Corporation, Santa Monica, CA, US.

Price Waterhouse Coopers (PwC) 2017, ‘Sizing the prize: What’s the real value of AI for your business and how can you capitalize?’, Price Waterhouse Coopers, London, UK.

Raso, F, Hilligoss, H, Krishnamurthy, V, Bavitz, C & Kim, L 2018, ‘Artificial Intelligence and human rights: Opportunities and risks”, 25 September. Berkman Klein Center for Internet and Society, Harvard University, Cambridge, MA, US.

Republic of South Korea 2017, *Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society,* Government of the Republic of South Korea, July 20.

Risse, M 2019, ‘Human rights and Artificial Intelligence: An urgently needed agenda’, *Human Rights Quarterly,* vol. 41, pp. 1-16.

Seeley, R 2019, ‘IDC: Annual worldwide spending on AI systems to reach $98 billion by 2023’, 5 September, International Data Corporation, Framingham, MA, US.

Stanford University 2016, ‘Artificial Intelligence and life in 2030’, viewed 27 August 2020, <https://ai100.stanford.edu/2016-report>.

State of Wisconsin v. Eric L. Loomis 2016, WI 68, 881 N.W. 2d 749.

Techworld 2019, ‘How tech giants are investing in artificial intelligence’, IDG Communication, Ltd., UK.

United Nations 2011, ‘Guiding principles on business and human rights: Implementing the United Nations “Protect, Respect and Remedy” framework’, endorsed by the Human Rights Council resolution 17/4, 16 June 2011.

Universal Declaration of Human Rights (UNDHR) 1948, U.N.G.A. Res. 217 A (III) (1948), 10 December.

US-China Economic and Security Review Commission Report to Congress 2019, 116th Congress (first session), November.

United States House Committee on Energy and Commerce, 116th Congress (2019-2020) 'H.R.2231 - Algorithmic Accountability Act of 2019, Referred to House Committee on Energy and Commerce', introduced10 April 2019.

Walz, A & Firth-Butterfield, K 2019, 'Implementing ethics into Artificial Intelligence: A contribution, from a legal perspective, to the development of an AI governance regime', *Duke Law & Technology Rev*iew, vol. 18, no. 1, pp. 180-231.

Wexler, R 2017, 'Code of silence: How companies hide software flaws that impact who goes to prison and who gets out', *Washington Monthly*, viewed 27 August 2020, <https://washington-monthly.com/magazine/junejulyaugust-2017/code-of-silence>.

The White House 2019, 'The National Artificial Intelligence Research and Development Strategic Plan: 2019', *Update: A Report by The Select Committee on Artificial Intelligence of the National Science & Technology Council*, June 2019, Executive Office of the President, Washington, D.C., US.

World Economic Forum (WEF) 2018, 'How to prevent discriminatory outcomes in machine learning', white paper, Global Future Council on Human Rights.

Yu, PK 2019, 'Intellectual property and human rights 2.0', *University of Richmond Law Review*, vol. 53, p. 1375.

Channel-PUFs: AI-Assisted Channel Estimation for Enhanced Wireless Network Security

C Lipps[1], SB Mallikarjun[2], M Strufe[1], HD Schotten[1,2]

[1]German Research Center for Artificial Intelligence
Intelligent Networks Research Group
Kaiserslautern, Germany

[2]University of Kaiserslautern
Division of Wireless Communications and Radio Positioning
Kaiserslautern, Germany

E-mail: Christoph.Lipps@dfki.de; Mallikarjun@eit.uni-kl.de; Mathias.Strufe@dfki.de; Schotten@eit.uni-kl.de

Abstract: *Next Generation Mobile Networks (NGMNs) are entering existing and future (industrial) wireless networks, associated with advantages such as higher data throughput, low latency, operation in almost real-time, and the microcell approach. However, this development also comes with drawbacks in the form of new attack vectors and security threats. Within this work, Physical Layer Security (PhySec) methods—the Channel-based Physically Unclonable Functions (PUFs)—are applied to derive symmetric cryptographic credentials and to establish a trustworthy sound, and secure communication between interacting entities. This is done by using a real-world implementation of an NGMN testbed to evaluate the adaptions to mobile radio. Artificial Intelligence (AI) in the form of the Linear Regression Algorithm (LRA) is applied to enhance the accuracy of the estimated channel profiles.*

Keywords: *Physical Layer Security; Physically Unclonable Functions; Secret Key Generation; Authentication; Plug & Trust; Channel-Estimation; Channel-PUFs; Wireless Network Security*

Introduction

Currently, the industrial landscape is going through a fundamental shift in the way communication takes place. The burst of traditional structures and communication links within the Industrial Automation and Control Systems (IACSs) is unmissable. There, a multitude of different sensors, actuators, and entities communicate with each other. The machines become smart and are capable of recognizing and independently deciding which processing step is to be carried out next. The environment as a whole is growing into an Industrial Internet of Things (IIoT), as well as into Cyber-Physical Production Systems (CPPSs).

The driving forces behind this development are the mobility, flexibility, and scalability of the connected devices and also, beyond that, the exchanged and transmitted data itself. As depicted in **Figure 1**, below, a wireless IIoT production environment can consist of a heterogeneous mixture of devices, such as, among others, Automated Guided Vehicles (AGVs), Robot Motion Control Systems (RMCSs), sensors, actuators, and even humans equipped and assisted by portable devices. As also illustrated, this can be achieved by the application of wireless communication with several independent Access Points (APs) and Radio Access Networks (RANs). The traditional and static plugs are simply not capable of providing the necessary features and handling the communication within the upcoming *Factories of the Future* and *Smart Factories* with application scenarios such as Machine-to-Machine (M2M) and Machine-to-Service (M2S) communication.

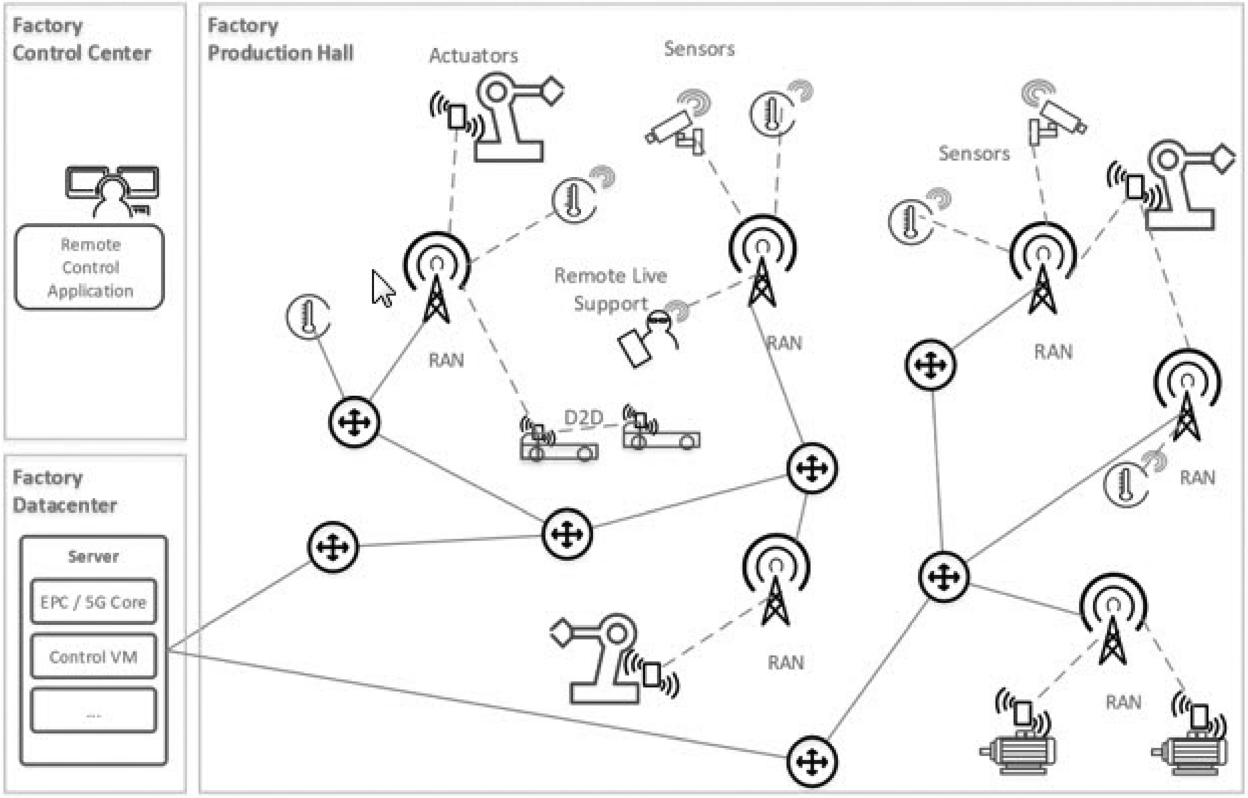


Figure 1: Wireless IIoT production environment

However, the usage of wireless communication involves drawbacks and security flaws. Due to the broadcast characteristic of wireless communication, the systems are often prone to miscellaneous cyberattacks such as eavesdropping or jamming. Furthermore, because of the open nature of wireless systems, it is important that only authorized and legitimate users can gain access to the networks. This aspect of network access is quite easier to monitor in traditional wired networks, quite simply because a "real" physical access is essential. Due to the characteristics mentioned in wireless networks, this monitoring is much more difficult and requires a certain amount of effort.

Besides the traditional IEEE802.11 WLAN, Next Generation Mobile Networks (NGMNs) (Roh *et al*. 2014) such as Long Term Evolution Advanced (LTE+) and the Fifth Generation (5G) establish themselves primarily on private networks, campus environments, and company premises. There, the regulations and directives by the legislator–the regulations with respect to frequency allocation and disturbing and jamming the public networks and infrastructures—are not that strict, and it is easier to set-up an intracompany network. Not only because of this, they are a promising alternative for future industrial networks. Furthermore, they offer communication technology related benefits such as almost real-time operation, low latency, and high data rates (Gundall *et al*. 2018).

Nevertheless, establishing security in wireless networks is still a crucial issue (Wang *et al.* 2013), and the aim of sound and secure communication remains, independent of the applied communication standard. In contrast to traditional and computational complex security mechanisms such as Certificates, Trusted Platform Modules (TPMs) (Gürgens *et al*. 2007), or pre-calculated cryptographic primitives which have to be brought in from the outside, Physical Layer Security (PhySec) offers benefits in terms of resource efficiency, scalability, and Perfect Forward Secrecy (PFS) (Menezes *et al*. 1996). Moreover, it is a worthwhile solution with respect to the independence of the wireless standard used, because the underlying operating principle is based on the physical properties of electromagnetic wave propagation and not on a specific communication standard.

Besides that, the possibilities offered by powerful and efficient Artificial Intelligence (AI) algorithms create new perspectives in IIoT security. For instance, the possibilities of Machine Learning (ML), in particular the Linear Regression Algorithm (LRA) presented in this work, enable an optimization of the system. Based on measured and network-current properties the *learning* is finding the parameters of the system (Kelleher 2019). More precisely, out of the three variants of ML algorithmss—supervised, unsupervised and reinforcement learning—, the LRA is assigned to superviesed learning (Boden 2016).

The remainder of this work is organized as follows. In Section 2 an introduction to Physical Layer Security is given to provide a basic introduction to the topic. Furthermore, a brief explanation about the wireless channel, Artificial Intelligence, and the capability of the enhanced Secret Key Generation (SKG) algorithm is given. Section 3 provides an overview of the applied testbed and the evaluated metrics. Building upon this, Section 4 describes the results and findings of the work. Finally, Section 5 concludes the work and provides an outlook on future work and the next steps to the further enhancements of the concept.

Physical Layer Security in Wireless Channels

As indicated in the previous Section, PhySec is a worthwhile, promising, and efficient possibility to face the cybersecurity requirements of IACSs, and to integrate security into existing and upcoming wireless networks. This section provides a brief introduction to the topic, as well as to the Discrete Cosine Transformation (DCT) and the Linear Regression Algorithm used to enhance the existing SKG algorithms.

Physical Layer Security as Enabler for Network Security

In general, the term Physical Layer Security comprises several different methods of how to utilize various physical characteristics of different materials and mediums (Lipps *et al*. 2019a). The most popular PhySec derivative are Physically Unclonable Functions (PUFs). Since they were introduced by Suh and Devadas (Suh & Devadas 2007), various principles have been developed to use. These are, besides others, electronic properties of components, manufacturing-related and uninfluenceable deviations of semiconductor circuits, such as Arbiter-PUFs, Ring-Oscillator PUFs (Suh & Devadas 2007) or SRAM-PUFs (Lipps *et al*. 2018), that can be summarized with the term *Silicon*-PUFs.

Relevant for this work are the so-called *Channel-PUFs* (Lipps, Duque Antón & Schotten 2019), that branch of PhySec which is utilizing the characteristics of electromagnetic waves and exploiting the randomness of a wireless channel. There are external influencing factors such as reflection, scattering, and diffraction that have an impact on the wireless channel. Additionally, according to the *principle of channel reciprocity* (Jakes 1974) the channel behaves similarly for two participating entities. For a third entity, for instance a potential attacker or eavesdropper, the channel has significantly different characteristics if it is distanced from the transmitter/receiver by a distance greater than half of the wavelength $\lambda/2.\lambda/2.$ This is the underlying premise of the PhySec SKG principle. However, the fundamental SKG building blocks, as indicated in **Figure 2**, are

- Channel Measurement,
- Quantisation,
- Information Reconciliation, and
- Privacy Amplification.

Generation of a Channel Profile

The basic step of SKG is the creation of a channel profile. As the principle is based on the values of the received signal strength, the channel variations are measured at both sides of the channel during the coherence time. This is the necessary step in order to learn about the present external channel influences. There are several different methods such as the Received Signal Strength Indicator (RSSI), Channel Impulse Response (CIR), Channel State Information (CSI), and especially in cellular environments the Referenced Signal Received Power (RSRP) to gain the information needed. Due to the principle of channel reciprocity these values are very similar on both sides of the radio channel.

Quantisation of the Channel Profile

Building up on the measured signals and to obtain a preliminary key the channel profile is quantised into vector bits (Proakis & Salehi 2002). Therefore, different methods are available. It can be done either on the whole sequence of the profile or on smaller blocks. Besides that, it can be distinguished between lossless quantisation, where all measured values are considered, and lossy quantisation, at which some values are dropped depending on the selected thresholds. To apply a multi-threshold quantisation, different methods can be used in order to select the threshold values. In this work, based on statistical values such as mean, standard deviation, and variance, the threshold values for different signal strength levels are calculated.

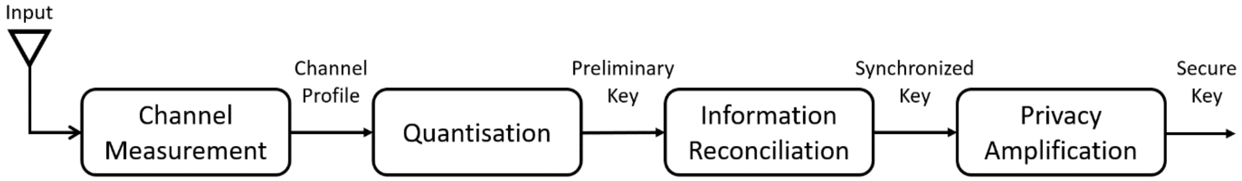


Figure 2: Building Blocks of the Secret Key Generation

Information Reconciliation

Due to errors occurring during the data transmission, such as superimposed noise, quantisation errors, and other disturbing signals, the preliminary keys are not completely identical on both sides of the channel. But, in order to use the keys for cryptographic operations, they must be the same; otherwise an encrypted data packet cannot be decrypted again. During the Information Reconciliation step, these errors are detected, localized, and corrected. This can be done, for instance, by using Turbo Codes.

Privacy Amplification

The Privacy Amplification step is an additional security feature of the SKG to achieve both, to minimize the amount of information a potential attacker can gain about the system, and to increase the entropy of the calculated key. This *Shannon Entropy* of a system is a measure of the unpredictability, disorder, and uncertainty within a system. Usually this is done by the application of a Secure Hashing Algorithm (SHA), such as SHA-3.

Besides this brief introduction to the SKG building blocks, a profound explanation and further descriptions of the various steps can be found in Ambekar & Schotten (2014), Zenger *et al*. (2015), Zenger *et al*. (2016), and Weinand *et al*. (2018). Furthermore, a survey of channel-reciprocity-based methods is given by Wang *et al*. (2015). An overview of the work of others, in respect to Physical Layer Security, Secret Key Generation, and Next Generation Mobile Network-PhySec, is also given in **Table 1**.

Scope of the work	**Work**
A general introduction of the PUF concept	Gassend *et al*. 2002
SRAM characteristics and suitability tests for PUFs	Lipps *et al*. 2018
A general PUF Introduction	Suh & Devadas 2007
Introduction to Secret Key Generation	Ambekar & Schotten 2014
WLAN PhySec to secure IoT environments	Zenger *et al*. 2015
PhySec with resource constrained and low-energy devices	Zenger *et al*. 2016
A security architecture for URLLC and a Plug & Trust protocol	Weinand *et al*. 2018
A description of a combination of PhySec and SDN	Lipps *et al*. 2019a
A survey about channel-reciprocity-based key establishment for wireless systems	Wang *et al*. 2015
WLAN based Testbed for wireless communication	Zhang *et al*. 2016
PhySec in Cellular Networks	Lipps *et al*., 2019b
Multi-Antenna Cellular Networks	Geraci *et al*. 2014
PhySec in mmWave Cellular Networks	Wang & Wang 2016

Table 1: Overview of Physical Layer Security work

Enhancing the Reciprocity of the Received Signals

According to the different building blocks of SKG, there are various starting points to enhance the existing algorithms. Besides approaches with respect to the *Information Theoretic* enhancements (Lin & Oggier 2014) and *Signal Processing* (Mukherjee *et al*. 2014), an improvement of the *Channel Reciprocity* is reasonable (Ambekar & Schotten 2014).

In this paper, another approach is presented to further increase the reciprocity of the channel. This is done by an adjusted prediction of the channel measures, based on the previous transmitted values of the channel. The approach is based on the one hand on the normalization of the measured channel values via the Discrete Cosine Transform, and on the other hand on the evaluation and learning of the channel characteristic by a Machine Learning algorithm, in particular the Linear Regression Algorithm with polynomial features.

The DCT can be expressed as a series of finite cosine functions oscillating at different frequencies. It is related to the *Fourier Transform* and similar to the *Discrete Fourier Transform (DFT)*, but using only real numbers. The mathematical representation of the DCT is given by

$$y(k) = \sqrt{\frac{2}{N-1}} \sum_{n=1}^{N} x(n) \frac{1}{\sqrt{1+\delta_{n1}+\delta_{nN}}} \frac{1}{\sqrt{1+\delta_{k1}+\delta_{kN}}} \cos\left(\frac{\pi}{(N-1)}(n-1)(k-1)\right)$$

where x represents the signal of the length N. $\delta_{ab}\delta_{ab}$ is the Kroenecker symbol (Trowbridge 1998), a function of two variables a and b.

In the context of this work a symmetric key of 128 bit length shall be generated. Therefore, a matrix Y of the size 128x128 is generated, from the image processing toolbox, which is, for instance, available in MATLAB.

$$Y = dctmtx(128) = \begin{bmatrix} x_{1x1} & x_{1x2} & \cdots & x_{1x128} \\ x_{2x1} & x_{2x2} & \cdots & x_{2x128} \\ x_{3x1} & x_{3x2} & \cdots & x_{3x128} \\ \vdots & \cdots & \ddots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ x_{128x1} & x_{128x2} & \cdots & x_{128x128} \end{bmatrix}$$

This matrix *Y* is represented as $T_{pq}T_{pq}$ in the formula below:

$$T_{pq} = \begin{cases} \dfrac{1}{\sqrt{M}}, & p = 0{,}0 \leq q \leq M-1 \\ \sqrt{\dfrac{2}{M}} \cos(\dfrac{\pi(2q+1)p}{2M}), & 1 \leq p \leq M-1, 0 \leq q \leq M-1 \end{cases}$$

The channel profile has 128 channel measurements, which is recorded for every second after the User Equipment (UE) (see below Section) is connected to the eNodeB. The measured channel profile is converted into 1x128 matrix X:

$$X = [x_{1x1}\ x_{1x2}\ \cdots\ x_{1x128}]$$

Furthermore, the profile *X* is multiplied with the DCT matrix *Y* to obtain *X'* which is the "new" channel profile:

$$X' = XY = [x'_{1x1}\ x'_{1x2}\ \cdots\ x'_{1x128}]$$

The main property of the DCT is to transform the majority of coefficients which represents energy sequence and on multiplying the DCT matrix with the channel profile, the spikes in the variation of the measured channel values are normalized which helps to generate a new enhanced channel profile *X'*.

In order to further increase the security level and the key strength, it is recommended to further increase the length and correspondingly the measurements and matrices. The 128 bits, in this case, are only a compromise between the generation time and the security level. In addition, the work focuses on the research of channel reciprocity and not on key length.

Artificial Intelligence, the Linear Regression Algorithm, and the Wireless Channel

With the increase in computing power and the ever-increasing efficiency of hardware, the options of using Artificial Intelligence are also growing. This also opens up new possibilities in the case of cybersecurity. But it still should be considered that there is neither *THE AI* nor is AI the *holy grail* of solving every problem of humankind. The topic of AI is very complex and must be divided into different areas which would go beyond the scope of this work. For a comprehensive and exhaustive introduction to this topic, please refer to Russel (2015).

Nevertheless, it is a very active field of research with activities in various domains, and even with respect to PhySec and the prediction of a wireless channel. For example, Ahrens *et al*. focus on the prediction of the time-variant transmission channels by using a simulation of Convolutional Neural Networks (CNN) (Ahrens *et al*. 2019). Hidden Markov Models (HMM) to predict the behaviour and spectrum of a channel based on the statistical Markov models are besides others, examined by Saad *et al*. (2016) and Eltom *et al*. (2015). Furthermore, Roy and Muralidahr (2015) try to predict the channel state in Cognitive Radio Networks with Hidden Markov Models. Another branch of ML, the Deep Learning with Artificial Neural Networks (ANNs), is used by Luo *et al*. for the prediction of the Channel State Information (2018). Sattiraju *et al*. are using ML and AI-assisted PhySec technologies for future wireless networks (2019). **Table 2** summarizes these works about AI based Channel Estimation.

Scope of the work	Work
Prediction of time-variant transmission channel with Convolutional Neural Networks	Ahrens *et al*. 2019
Hidden Markov Models to predict the channel behaviour	Saad *et al*. 2016
Hidden Markov Models for spectrum prediction	Eltom *et al*. 2015
Hidden Markov Models for Cognitive Radio Network prediction	Roy & Muralidhar 2015
Prediction of the Channel State Information	Luo *et al*. 2018
Machine Learning and AI-assisted PhySec for future wireless networks	Sattiraju *et al*. 2019
General introduction to Artificial Intelligence	Russel 2015
Introduction to Machine Learning and its subcategories	Alpaydin 2016

Table 2: Overview of work about Artificial Intelligence and Channel Prediction

Besides the AI research, there is some work on PhySec in cellular networks, such as Geraci *et al*., dealing with PhySec in Multi-Antenna cellular networks (Geraci *et al*. 2014). Yang *et al*. (2015) intend to safeguard the upcoming 5G wireless communication with PhySec, and Chen and Willems (2018) provide a stochastic geometry model of networks.

The LR algorithm (Freedman 2009) as treated in this paper, one of the supervised Machine Learning Algorithms, works based on co-relating dependent and independent variables. The representation of a linear equation is the combination of a set of input values XX and a predicted set of output values Y. A coefficient scale factor $\beta_0\beta_0$ and a bias coefficient $\beta_1\beta_1$ are assigned to a linear equation:

$$Y = \beta_0 + \beta_1 X$$

Based on the Mean Square Error value, the best fit for β_0 and β_1 are decided. To get the best prediction or best fit, a cost function JJ, as sort of a minimization problem that tries to minimize the error between the predicted and the actual value, is used:

$$J = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2$$

The data distribution in the channel profile is complex and so the linear regression will be in the under-fit state, as it cannot cover all the pattern in data, so the complexity of the model is increased by adding to a higher power to the original feature,

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_n X^n$$

adding as new features and hence overcoming under-fit nature. The main idea behind using the AI-based method is to enhance the reciprocity of the channel profile by producing a similar channel profile on both entities (eNodeB and UE; see section below). The designed AI model is trained with channel measurements of eNodeB, UE, and the average of both the entities to predict the enhanced channel profile for both the entities (Russell 2011).

The Bit Disagreement Rate

In order to use the SKG resulting keys for cryptographic operations—encryption and decryption of data—it is mandatory to create identical keys on both sides of the channel. One measure for the matching between two sequences of equal length is the Bit Disagreement Rate (BDR) or Hamming Distance $\delta\delta$ (Robinson 2003). It expresses the number of positions at which the corresponding bits are different. Mathematically this can be described as with the sum $\Sigma\Sigma$ of a finite alphabet as well as $x = (x_1, \ldots, x_n)x = (x_1, \ldots, x_n)$ and y= $(y_1, \ldots, y_n)$= $(y_1, \ldots, y_n)$ two symbols of the length nn of $\Sigma^n.\Sigma^n$. Therefore, the hamming distance is given by

$$\delta(x, y) \coloneqq \left|\{j \in \{1, \ldots n\} \mid x_j \neq y_j\}\right|$$

If this $\delta\delta$ is too large, it is not feasible to correct the preliminary key in the information reconciliation step. Therefore, the BDR or $\delta\delta$ is a good metric for the quality of the SKG.

Cellular-PhySec Testbed

As mentioned in the previous section, a few works are dealing with PhySec and with AI methods to enhance the Physical Layer Security approach. Nevertheless, most of the work is theoretical, simulations and models. In order to gain real-world results a testbed is set up. Because there is no 5G equipment available yet, the described testbed is based on a Software Defined Radio (SDR) Long Term Evolution (LTE) Base Station (BS). As soon as 5G user equipment, core functionality, and radio base stations are available, the testbed will be extended accordingly.

Long Term Evolution-Based Testbed

The main components of the testbed are depicted in **Figure 3**. It consists of the open source software srsLTE, developed by Software Radio Systems (Software Radio Systems Limited 2019) under AGPLv3. It implements LTE Release 8 and provides an entire Software Defined Radio LTE User Equipment (UE). Furthermore, it provides a complete eNodeB and lightweight LTE Evolved Packet Core (EPC) network implementation, including the functionalities of a Serving Gateway (S-GW), Packet Data Network (PDN)-Gateway, Home Subscriber Server (HSS), and Mobility Management Entity (MME). It is possible to run the system on a generic MiniPC connected to a Universal Software Radio Peripheral (USRP) or any other SDR supported by srsLTE software. A commercial LTE dongle is used as UE. The radio channel between eNodeB and UE is set up as the basis for the analysis.

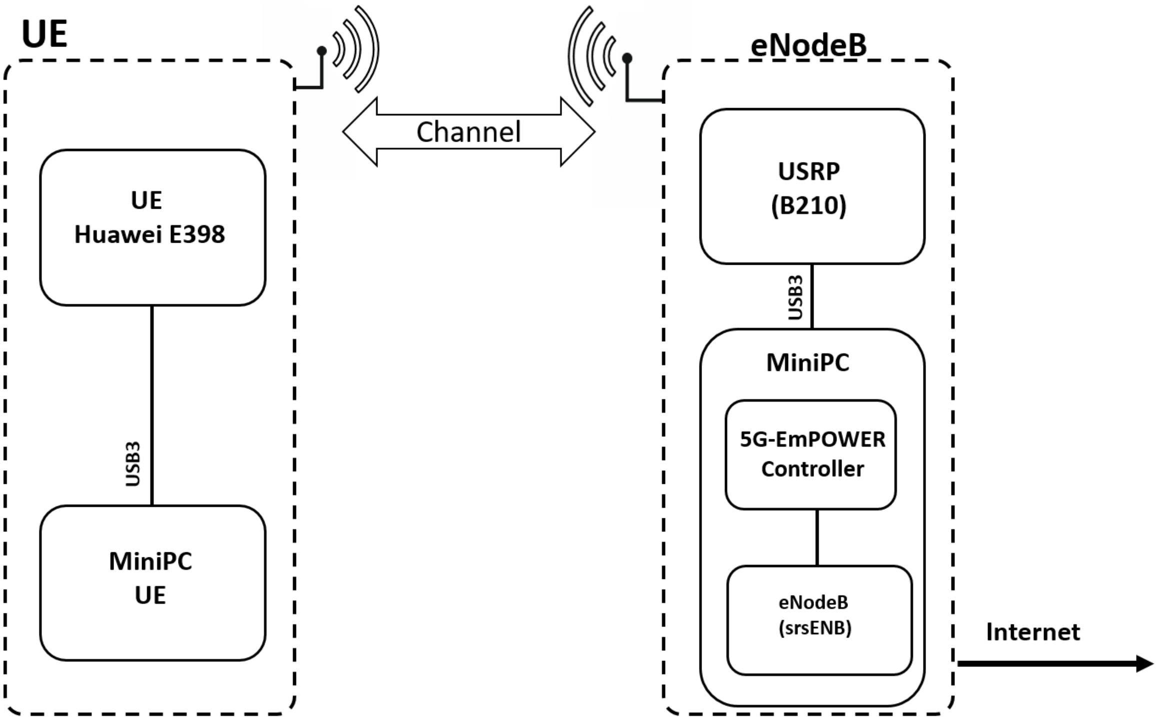


Figure 3: LTE-Testbed for Validation

Due to legal regulations, the transmission power of the LTE testbed is limited to avoid interferences and jamming of existing public radio and cellular networks. The setup is built in a normal laboratory and office environment of about 4x4 meters, with standard office equipment. Transmitter and receiver are statically positioned as a distance of 2 meters. Interferences arise through parallel networks and reflections through walls, office equipment, and devices. This setup provides the framework for the measurements of reference values with which the LR algorithm is trained. Besides that static set-up, the evaluation is extended by a mobile scenario. In order to get results with varying channel characteristics, a mobile variation is included into the set-up. Therefore, one of the nodes is still at a fixed position, but the second node is plugged on top of a mobile vacuum cleaner robot. This is done to integrate motion and varying effects to the channel: scattering, refection, and detraction of radio signals. The vacuum cleaner is moving randomly in the office space with another benefit: the vacuum cleaner battery is supplying the node, too.

A further and more detailed description of the testbed as well as implementation information is available in Lipps, Duque Antón and Schotten (2019) and Lipps *et al*. (2019a).

Results and Discussion

After the basic principles of PhySec, SKG, and LR were introduced in the previous sections, the testbed and the value of BDR has been described. Within this work, only the first step of the SKG has been implemented (see **Figure 2**) because the focus of the work is the enhancement of the channel Profile.

As described in section 2, different quantisation methods are used. The channel profiles are converted to 2 bits in the *Standard Deviation Quantisation, Standard Deviation Block Quantisation, Variance Quantisation, and Variance Block Quantisation*; therefore, the generated key—and the reference value for the BDR—is 256 bits. For the *Mean Quantisation, Mean Block Quantisation, Median Quantisation and Median Block Quantisation*, the key length is 128 bits. This shall be considered when comparing the BDR values in **Tables 3-5**. The absolute value in percent gives a better overview at this point. Furthermore, the BDR values in **Tables 3-5** represent the average value of 10 independent measurements, but with respect to the different quantisation methods.

Evaluation of the static testbed

The results of the evaluation are separated into three different parts. First of all, a non-enhanced measurement of the applied wireless channel is prepared. This is done in order to generate a comparative value of the channel and to monitor the enhancements of the DCT and LR-based improvements.

Non-Enhanced SKG method

First of all, the pure channel measurement is done, without any form of error correction, reciprocity enhancement, and further adjustments. As summarized in **Table 3**, the BDRs with the different quantisation methods are quite high, especially with the *Mean Quantisation* and the *Mean Block Quantisation* whereas the BDR is 50% and 54%, respectively. In other words, on average every second bit is disagreeing to the compared bit sequence. Even the other quantisation methods perform not much better. However, the *Median Block Quantisation* is the best one, but there are also at least 30% different bits.

DCT normalized SKG method

To compensate for the drawbacks and to lower the BDR, a channel enhancement is done by using the DCT normalization. When compared the results with the non-enhanced version, the BDR for all different quantisation methods is reduced significantly. As shown in Table 4 the BDR is significantly lower than in Table 3. The BDR is of a level between 3.906% (Standard Deviation Block Quantisation) and 13.672% (Variance Block Quantisation) and thereby at least one-tenth of the non-enhance BDR.

Environment	Reciprocity Enhancing Method	Quantisation Method	Number of Disagreed Bits	BDR (%)
Static Environment	none	Standard Deviation Quantisation	94	36.719
		Standard Deviation Block Quantisation	90	35.156
		Variance Quantisation	93	36.719
		Variance Block Quantisation	84	32.812
		Mean Quantisation	64	50
		Mean Block Quantisation	70	54.688
		Median Quantisation	63	49.219
		Median Block Quantisation	39	30.469

Table 3: Non-Enhanced SKG method's BDR in static environment

Environment	Reciprocity Enhancing Method	Quantisation Method	Number of Disagreed Bits	BDR (%)
Static Environment	DCT normalization	Standard Deviation Quantisation	6	2.344
		Standard Deviation Block Quantisation	10	3.906
		Variance Quantisation	10	3.906
		Variance Block Quantisation	35	13.672
		Mean Quantisation	6	4.688
		Mean Block Quantisation	8	6.250
		Median Quantisation	10	7.812
		Median Block Quantisation	64	50

Table 4: DCT normalized SKG method's BDR in static environment

ML-based with DCT normalized SKG method

At least, the LR algorithm is applied in addition to the DCT normalization. The level of the BDR values decreases, again—especially the BDR of the *Mean Quantisation*, which had the worst effects in the non-enhanced method, there is only one single disagreeing bit. This very low BDR, see **Table 5**, can also be applied for the *Variance Quantisation* (0.391%) and the *Standard Deviation Quantisation* (0.781%). In general, the ML-based DCT normalized SKG have, except the *Standard Deviation Block Quantisation* (23 disagreeing bits), a BDR of less than 10 bits.

Evaluation of the mobile testbed

Besides the set-up with two static sending/transmitting nodes, the behaviour of the channel is investigated with one static and one mobile node. This extension enables results that are more similar to IIoT environments. Due to the movement, there are more channel variations but there is higher entropy in the channel, too. Subsequently, the same evaluations are made as in the static environment.

Environment	Reciprocity Enhancing Method	Quantisation Method	Number of Disagreed Bits	BDR (%)
Static Environment	ML-based with DCT normalization	Standard Deviation Quantisation	2	0.781
		Standard Deviation Block Quantisation	23	8.984
		Variance Quantisation	1	0.391
		Variance Block Quantisation	7	2.734
		Mean Quantisation	1	0.781
		Mean Block Quantisation	4	3.125
		Median Quantisation	4	3.125
		Median Block Quantisation	6	4.688

Table 5: ML-based with DCT normalized SKG method's BDR in static environment

DCT normalized SKG method

In the second step, the DCT normalization of the SKG values, and to enhance the SKG method, a DCT normalization is used. **Table 7** gives an overview of significantly decreased BDR values compared to the non-enhanced method in **Table 6**. In this case, for instance the *Standard Deviation Quantisation* (8.203%) and the *Standard Deviation Block Quantisation* (7.031%) show deviating bits below 10%. Just with the *Mean Block, Median*, and *Median Block Quantisation* about every fifth bit is disagreeing.

Environment	Reciprocity Enhancing Method	Quantisation Method	Number of Disagreed Bits	BDR (%)
Mobile Environment	none	Standard Deviation Quantisation	85	33.547
		Standard Deviation Block Quantisation	91	35.547
		Variance Quantisation	58	22.656
		Variance Block Quantisation	91	35.547
		Mean Quantisation	45	35.156
		Mean Block Quantisation	59	46.094
		Median Quantisation	43	33.594
		Median Block Quantisation	59	46.094

Table 6: Non-Enhanced SKG method's BDR in mobile environment

Environment	Reciprocity Enhancing Method	Quantisation Method	Number of Disagreed Bits	BDR (%)
Mobile Environment	DCT normalization	Standard Deviation Quantisation	21	8.203
		Standard Deviation Block Quantisation	18	7.031
		Variance Quantisation	24	9.375
		Variance Block Quantisation	41	16.016
		Mean Quantisation	21	16.406
		Mean Block Quantisation	28	21.875
		Median Quantisation	24	18.750
		Median Block Quantisation	26	20.313

Table 7: DCT normalized SKG method's BDR in mobile environment

Non-Enhanced SKG method

As in the previous section, the results in the mobile environment are measured first to get the "pure" non-enhanced values. As summarized in **Table 6** and as with the results of the static environment, the BDRs are quite high. For instance with the *Mean Block Quantisation* (46.094%) and the *Median Block Quantisation* (46.094%), about half of the measured bits are disagreeing.

ML-based with DCT normalized SKG method

Table 8, below, summarizes the values of the "full" LRA and DCT enhanced results of the mobile scenario. Here too, as in the static case, the results are significantly better than without the extensions, although not quite as outstanding as in **Table 4**. It is the *Standard Deviation* and the *Variance Quantisation* with 4 disagreeing bits, respectively. This is too, a very good result for a set-up of an ever-changing and varying environment, but not yet sufficient for an application in a cryptographic system. However, it should be considered that it is merely the first step of the SKG procedure and thus provides a good and solid basis for the subsequent steps.

Environment	Reciprocity Enhancing Method	Quantisation Method	Number of Disagreed Bits	BDR (%)
Mobile Environment	ML-based with DCT normalization	Standard Deviation Quantisation	4	3.125
		Standard Deviation Block Quantisation	19	7.422
		Variance Quantisation	4	3.125
		Variance Block Quantisation	32	15.500
		Mean Quantisation	8	6.250
		Mean Block Quantisation	11	8.594
		Median Quantisation	10	7.812
		Median Block Quantisation	8	6.250

Table 8: ML-based with DCT normalized SKG method's BDR in mobile environment

Comparison of the individual methods

As already indicated in the previous sections, the applied enhancements perform very well. **Figures 4 and 5**, below, illustrate this impressively. On the y-axis the BDR in percent is given, whereby the x-axis shows quantisation method. The yellow curve, with the small triangles, represents the non-enhanced channel profile; the red curve with the diamonds represents the DCT normalized curve; and the blue line with the squares is the AI+DCT values. Noteworthy is that for the non-enhanced versions, the BDR is, with one outlier, always above 30%. In contrast to this, the DCT normalized curve and the AI+DCT values, perform much better and show BDRs of, in most cases, less than 10%. This is a significant minimisation of the BDR. In most cases there is nearly no error anymore.

Besides that, **Figure 5** indicates that the DCT normalization is not as effective in mobile environments as in the static set-up, but in combination with the LRA almost reaches the level of the static values.

Physical Layer Security is not a completely new concept. There is some work of others dealing with this topic. However, on the one hand, the existing approaches are related to PhySec in IEEE802.11 WLAN; and on the other hand, most of the work is theoretical. Results in this area prove the general applicability and performance of the algorithms. Nevertheless, for several reasons, it remains an active field of research with potential for improvement.

As already mentioned, there is a lot of theoretical work. An implementation in real-world leads in many cases to deviating results. But it is mandatory to use real-world measured values and to make improvements on this data. Besides that, the rise of Next Generation Mobile Networks and their impact on the ways of communication opens up new requirements in case of securing the networks. Because of this, it is necessary to put effort in the development, adaption, and enhancement of security mechanisms such as PhySec.

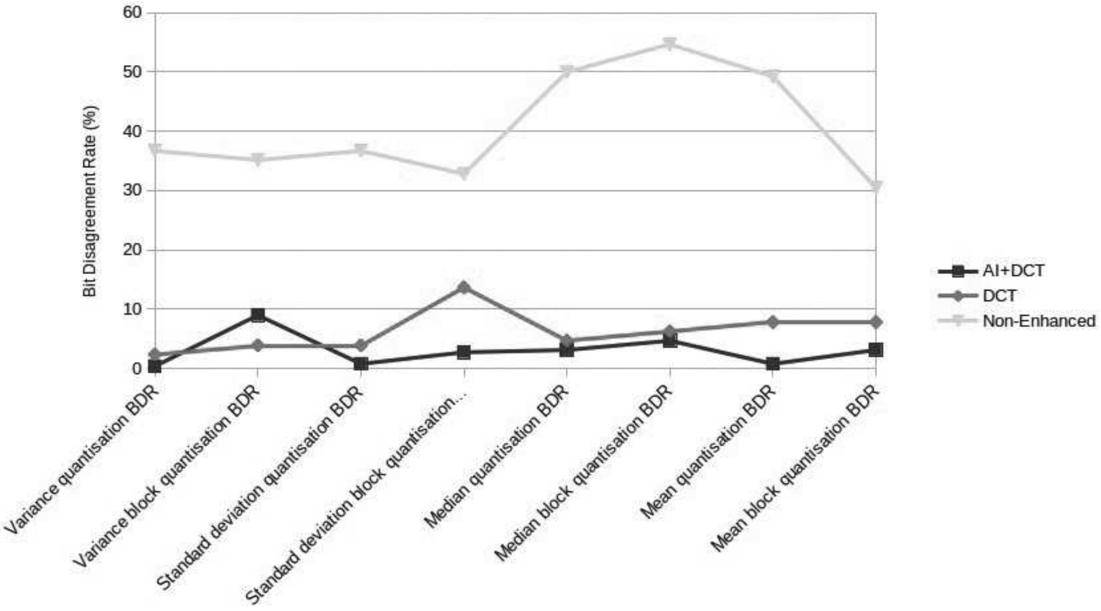


Figure 4: Bit Disagreement Rate in a static environment

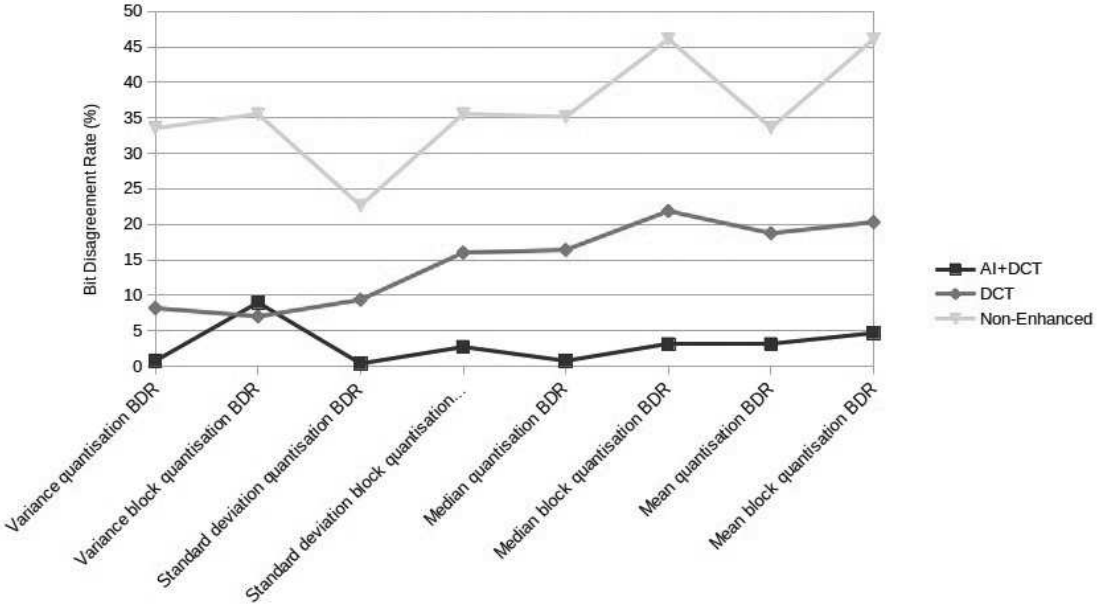


Figure 5: Bit Disagreement Rate in a mobile environment

The results of this work validate the possibilities and the potential inherent in the algorithms. Just by applying the DCT normalization and a Machine Learning algorithm, the researchers were able to minimize the Bit Disagreement Rate significantly. Through the use of even more training data under various conditions the quality of the ML can be further improved.

Conclusion, Outlook, and Next Steps

There is an ongoing fundamental change in the way communication and data transmission are taking place, especially in industrial environments. Most of the future networks will contain a multitude of wireless connected devices and entities. Nevertheless, securing these networks is a crucial and costly task.

Physical Layer Security algorithms, especially Channel-PUFs, offer a worthwhile and promising solution to the requirements of (wireless) network security. Instead of integrating pre-calculated cryptographic credentials and TPMs from the outside, inherent given information is used to establish security.

The results of this work already show a very good performance of the Machine Learning and Linear Regression algorithms. Compared to the non-enhanced values, the performance of the Bit Disagreement Rate is significantly better with the Reciprocity Enhancement by the Discrete Cosine Transformation and the Linear Regression algorithm.

Nevertheless, the tests included just static connections. To gain more real-world-like results, mobile devices must be integrated into the test, too. These mobile scenarios fit much better to the circumstances in industrial use-cases. Most of the work of others contains theoretical approaches and simulations. The results of this testbed evaluation should be compared to the theoretical works to make a better statement about the quality of the enhanced approaches.

Furthermore, there are multiple different Machine Learning and Artificial Intelligence methods, such as Reinforcement Learning, Neural Networks, Support Vector Machines, or Hidden Markov Models, that must be considered and validated in real-world scenarios. This is necessary in order to compare them with each other, regarding for instance the energy consumption for the algorithms, the speed and the accuracy of the algorithms.

Acknowledgements

This paper is an extension of work originally presented in Lipps, C, Mallikarjun, SB, Strufe, M & Schotten, HD 2020, 'Enhanced Channel-PUFs: The Advantages of Artificial Intelligence for Channel Estimation', *Proceedings of the International Conference on Cyber Warfare and Security (ICCWS-2020)*, Norfolk, VA, US, DOI: 10.34190/ICCWS.20.044.

References

Ahrens, J, Ahrens, L & Schotten, HD 2019, 'A Machine Learning Method for Prediction of Multipath Channels'*, Electrical Engineering and System Science, Signal Processing, arXiv:1909.04824v1.*

Alpaydin, E 2016, *Machine learning,* Massachusetts Institure of Technology Press, Cambridge, MA, US.

Ambekar, A & Schotten, HD 2014, 'Enhancing Channel Reciprocity for effective Key Management in Wireless ad-hoc Networks'*, IEEE 79th Vehicular Technology Conference (VTC Spring)*, Seoul, KR.

Boden, MA 2016, *AI—Its nature and future,* 1st edn., Oxford University Press, Oxford, UK.

Chen, B & Willems, FMJ 2018, 'Secret Key Generation over Biased Physical Unclonable Functions with Polar Codes', *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 435-45, DOI: 10.1109/JIOT.2018.2864594.

Eltom, H, Kandeepan, S, Moran, B & Evans, RJ 2015, 'Spectrum occupancy prediction using a Hidden Markov Model', *9th International Conference on Signal Processing and Communication Systems (ICSPCS),* Cairns, QLD, Australia, DOI: 10.1109/ICSPCS.2015.7391772.

Freedman, DA 2009. *Statistical Models: Theory And Practice,* 2nd edn., Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town.

Geraci, G, Dhillon, HS, Andrews, JG, Yuan, J & Collins, IB 2014, 'Physical Layer Security in Downlink Multi-Antenna Cellular Networks', *IEEE Transactions on Communications,* vol. 62, no. 6, pp. 2006-21, DOI: 10.1109/TCOMM.2014.2314664.

Gundall, M, Schneider, J, Schotten, HD, Aleksy, M, Schulz, D, Franchi, N, Schwarzenberh, N, Markwart, C, Halfmann, R, Rost, P, Wübben, D, Neumann, A, Düngen, M, Neugebauer, T, Blunk, R, Kus, M & Grießbach, J 2018, '5G as Enabler for Industrie 4.0 Use Cases: Challenges and Concepts'*, IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, Turin, IT, DOI: 10.1109/ETFA.2018.8502649.

Gürgens, S, Rudoplh, C, Schermann, D, Atts, M & Plaga, R 2007, 'Security Evaluation of Scenarios Based on the TCG's TPM Specification', In: J, Biskup & J, Lopez (eds.), *Computer Security -*

ESORICS 2007, Lecture Notes in Computer Science, vol. 4743. Springer, Berlin Heidelberg, DE.
Jakes, WC 1974, *Microwave Mobile Communication.*Wiley-IEEE Press.

Kelleher, JD, 2019, *Deep Learning.* Massachusetts Institure of Technology Press, Cambridge, MA, US.

Lin, & Oggier, F 2014, 'Coding for Wiretap Channel', *Physical Layer Security in Wireless Communications*, eds. X. Zhou, L, Song & Y, Zhang,. pp. 17-32, CRC Press, New York, US.
Lipps, C, Duque Antón, S & Schotten, HD 2019, 'Enabling Trust in IIoT: A PhySec based Approach', *14th International Conference on Cyber Warfare and Security (ICCWS2019)*, Stellenbosch, ZA, *Proceedings of the 14th International Conference on Cyber Warfare and Security*.

——, Strufe, M, Mallikarjun, SB & Schotten, HD 2019a, 'Physical Layer Security for IIoT and CPPS: A Cellular-Network Security Approach', *Mobilkommunikation—Technologien und Anwendungen (ITG),* vol 288, pp. 82-86, Osnabrück, DE.

——, Strufe, M, Mallikarjun, SB & Schotten, HD 2019b, 'PhySec in Cellular Networks: Enhancing Security in the IIoT'*, European Conference on Cyber Warfare and Security (ECCWS-2019)* Coimbra, PT.

——, Weinand, A, Krummacker, D, Fischer, C & Schotten, HD 2018, 'Proof of concept for IoT device authentication based on SRAM PUFs using ATMEGA 2560-MCU'*, 1st International Conference on Data Intelligence and Security (ICDIS),* South Padre Island, Texas, US.

Luo, C, Ji, J, Wang, Q, Chen, X & Li, P, 2018, 'Channel State Information Prediction for 5G Wireless Communications: A Deep Learning Approach', IEEE Transactions on Network Science and Engineering, vol. 7, no. 1, pp. 227-36, DOI: 10.1109/TNSE.2018.2848960.

Menezes, AJ, van Oorschot, PC & Vanstone, SA 1996, *Handbook of Applied Cryptography,* CRC Press Series on Discrete Mathematics and Its Applications, Middletown, New Jersey, US.

Mukherjee, A, Fakoorian, SAA, Huang, J & Swindlehurst, L 2014, 'MIMO Signal Processing Algorithms for Enhanced Physical Layer Security', In: X. Zhou, L. Song & Y. Zhang, eds. *Physical Layer Security in Wireless Communications,* pp. 93-114, CNC Press, New York, US.

Proakis, JG & Salehi, M 2002, *Communication Systems Engineering.* 2nd edn, Prentice-Hall Inc., New Jersey, US.

Robinson, DJS 2003. *An Introduction to Abstract Algebra,* 2nd edn., De Gruyter Textbook, Berlin, DE.

Roh, W, Seol, J-Y, Park, J, Lee, B, Lee, J, Kim, Y, Cho, J, Cheun, K & Aryanfar, F 2014, 'Millimeter-wave Beamforming as an enabling Technology for 5G Cellular Communications: Theoretical feasibility and prototype results', *IEEE Communications Magazine,*vol. 52, no. 2, pp. 106-13, DOI: 10.1109/MCOM.2014.6736750.

Roy, PP & Muralidhar, M 2015, 'Hidden Markov Model based Channel State Prediction in Cognitive Radio Networks', *International Journal of Engineering Research & Technology (IJERT), vol.* 4, no. 2, pp. 391-4.

Russell, TW 2011, ‘Beyond Multiple Regression: Using Commonality Analysis to Better Understand R2 Results’,*Gifted Child Quarterly*, vol. 55, no. 4, pp. 313-8.

Russel, S 2015, *Artificial Intelligence: A Modern Approach, Global Edition.* 3rd edn., Pearson Education Limited, Harlow, US.

Saad, A, Staehle, B & Knorr, R 2016, ‘Spectrum prediction using hidden Markov models for industrial cognitive radio’, *12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob),* New York, NY, US, DOI: 10.1109/WiMOB.2016.7763231.

Sattiraju, R, Weinand, A & Schotten, HD 2019, ‘AI-assisted PHY technologies for 6G and beyond wireless networks’*, The 1st 6G WIRELESS SUMMIT,* Levi, FI.

Software Radio Systems Limited 2019, *Sofware Radio Systems,* Digital Bond, viewed 09/2019, <https://www.softwareradiosystems.com>.

Suh, GE & Devadas, S 2007, ‘Physical Unclonable Functions for Device Authentication and Secret Key Generation’, *44th ACM/IEEE Design Automation Conference*, San Diego, CA, US.
Trowbridge, JH 1998, ‘On a Technique for Measurement of Turbulent Shear Stress in the Presence of Surface Waves’, *Journal of Atmospheric and Oceanic Technology,* vol. 15 no. 1, pp. 290-8.

Wang, C & Wang, HM 2016, ‘Physical Layer Security in Millimeter Wave Cellular Networks’ *IEEE Transactions on Wireless Communications,* vol. 15, no. 8, pp. 5569-85.

Wang, H, Zhou, X & Reed, MC 2013, ‘On the Physical Layer Security in Large Scale Cellular Networks’, *IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, CN.

Wang, T, Liu, Y & Vasilakos, AV 2015, ‘Survey on channel reciprocity based key establishment techniques for wireless systems’, *Wireless Networks,* vol. 21, pp. 1835-1846, Springer, US.

Weinand, A, Karrenbauer, M & Schotten, HD 2018, ‘Security Solutions for Local Wireless Networks in Control Applications based on Physical Layer Security’, *IFAC-PapersOnLine*, vol. 51, no. 10, pp. 32–9.

Yang, N, Wang, L, Geraci, G, Elkashlan, M, Yuan, J & Di Renzo, M 2015, ‘Safeguarding 5G wireless communication networks using physical layer security’, *IEEE Communications Magazine,* vol. 53, no. 4, pp. 2027.

Zenger, CT, Pietersz, M Zimmer, J, Posielek, JF, Lenze, T & Paar, C 2016, ‘Authenticated key establishment for low-resource devices exploiting correlated random channels’, *Computer Networks,* vol. 109, no. 1, pp. 105-23.

——, Zimmer, J, Pietersz, M, Posielek, JF & Paar, C 2015, ‘Exploiting the Pyhsical Environment for Securing the Internet of Things’*, Proceedings of the 2015 New Security Paradigms Workshop*, Twente, NL.

Zhang, J, Woods R, Duong, TQ, Marsahll, A, Ding, Y, Huang, Y & Xu, Q 2016, ‘Experimental Study on Key Generation for Physical Layer Security in Wireless Communication’, *IEEE Access,* no. 4, pp. 4464-77.

Enlisting Propaganda for Agenda Building: The Case of the Global Times

DS Wilbur

School of Journalism
The University of Texas at San Antonio
San Antonio, Texas, United States

Email: Douglas_wilbur@yahoo.com

Abstract: *This article uses agenda building theory to examine how the People's Republic of China is using propaganda in the form of news to build an agenda within global news media as part of their Three Warfares Strategy. This qualitative content analysis of the English language version of the* Global Times *revealed that the PRC is building an offensive news agenda to directly challenge the credibility and legitimacy of its rivals. It is obsessively preoccupied by receiving blame for COVID-19-related mistakes and threats of economic decoupling.*

Keywords: *Propaganda, Agenda Building Theory, Three Warfares Strategy, Inter-Media Agenda Setting, China, Global Times*

Introduction

The People's Republic of China (PRC) is intensifying its global propaganda campaign, including accusations that the U.S. Army created the COVID-19 virus (Yuan 2020). Nearly half of Americans reported being exposed to disinformation about COVID-19 (Schaeffer 2020), and some of this is undoubtedly PRC propaganda. However, COVID-19 is only one prong of a larger campaign of information warfare designed to facilitate PRC's rise to superpower status. The PRC's Three Warfares concept specifically calls for strategic psychological operations, the infiltration and subversion of news media, and exploitation of adversary legal systems (Livermore 2018). While manipulation of adversarial public opinion is not a new concept, it is dangerous to the sovereignty of Western democratic nations with open societies.

Only two scholarly articles exist on the Three Warfares. The PRC made aggressive but ultimately unsuccessful attempts to use news media to undermine India's political will in a conflict over Bhutan (Bisht, Jain & Gambhir 2019). The U.S. has apparently been less successful in countering the effects of Three Warfares as evidenced by the lack of an effective response to aggressive cyberwarfare (Iasiello 2016). Given the obvious dangers of the Three Warfares Strategy to the integrity of Western journalism, there is scant scientific literature about the threat it poses. In democratic nations, a journalist's autonomy and independence from manipulative forces is a core professional value and necessity for a healthy democratic society (Lauk & Harro-Loit 2017).

In a broad geo-political context, if the PRC can subvert the internal discourse of open Western societies and steer that debate in a manner that is advantageous to its foreign policy goals, this would represent a coup for the Three Warfares Strategy. For instance, influencing the populations of some

ISSN 1445-3312 Print/ISSN 1445-3347 Online

nations like the U.S. and Australia to withhold military support for the defence of Taiwan would render a PRC invasion of the island more likely to succeed. Fortunately, it is possible to deduce what the PRC's goals for the Three Warfares Strategy are by scientifically examining the content of what their propaganda is actually saying and attempting to do. Discovering their ostensible goals creates the opportunity for Western open societies to defend themselves against subversive influence.

One established method of influencing the news media agenda and framing of issues is described by the agenda-building theory. The theory posits that the news media's agenda is the result of diverse groups in society that exert influence over journalists' work by exploiting their limited resources through offering information subsidies. These information subsidies are controlled, and deliberate information packages given to a recipient at negligible cost and effort (Lariscy, Avery & Sohn 2010). These subsidies help to transfer salience over what topics are important to cover as well as what features and attributes of those topics are important (Lee & Riffe 2017). The PRC's most logical method of influencing the Western news media is by using scientifically proven methods like agenda building. This begs the question, what agenda is the PRC trying to build within the U.S. news media? The purpose of this study is to ascertain what agenda the PRC is trying to build in order to provide a starting point for future research.

Review of Literature

Propaganda

Propaganda can be defined as the "deliberate, systematic attempt to shape perceptions, manipulate cognitions, and direct behavior to achieve a response that furthers the desired intent of the propagandist" (Jowett & O'Donnell 2015, p. 7). All types of propaganda are forms of inherently manipulative strategic communication. Propaganda is always deliberately created in order to influence human behaviour that benefits the propagandist (Wilbur 2017). Propaganda is inseparable from ideology due to its usage by actors to gain or maintain power structures (Burnett 1989). Ideology prescribes for the propagandist a set of beliefs, values, attitudes, and behaviours that the audience should adopt as the end result (Jowett & O'Donnell 2015).

In China, propaganda is not viewed pejoratively but is rather an ancient cultural tradition that is accepted as normal and necessary. It reflects a deeply collectivist society where communication from those in authority creates harmony in society when the population adheres to it (Lin 2017). In many cases, news is synonymous with propaganda, which is both appropriate and acceptable (Wang, Sparks & Yu 2018). Even the PRC news media outlets, like the *Global Times*, specifically aim overtly nationalistic and propagandistic content at Western audiences (Zeng & Sparks 2011). The current General Secretary, Xi Jinping, has greatly intensified domestic and foreign propaganda, even tapping into ancient Confucian ideals in a more flexible manner than his predecessors (Chang & Ren 2018). PRC propaganda is significantly influenced by its military strategy to the point where its overarching propaganda plan creates harmony between military, diplomatic, economic, and other goals (Pei-Ling 2014). Clearly, the PRC takes its propaganda very seriously, and academic research should as well.

Agenda Building

Agenda-building theory is optimal for this study since it specifically examines how people, organizations, and coalitions attempt to directly influence what the news media discusses and how it is framed. Since the second element of the Three Warfares Strategy is to infiltrate and subvert adversary news media, this describes agenda building perfectly. It developed from agenda-setting theory, which holds that the media sets the national agenda by telling the public what issues are important. Any topic that is on the national news agenda will be discussed more frequently and thought about more saliently in people's minds (McCombs 2014). However, the media's news agenda is itself influenced by others through the process of salience formation in a process of reciprocal and shared influence among a variety of stakeholder groups and constituencies (Hughes & Dann 2009). Journalists need information to report on that non-journalists have, which is offered to journalists as a form of information subsidy. The PRC state-run news outlets, like *Xinhua*, have been proven to be a source of information subsidies for Western journalists, even though the outlets' propaganda role has been identified. Additionally, journalists have relationships with people and institutions who often provide these subsidies, and these relations are proven to impact their work (Kiousis *et al*. 2016).

News media outlets can build an agenda for other news media outlets in what is known as inter-media agenda building. Journalists are well known to examine and use the work of other journalists when developing stories (Denham 2010). *The New York Times*, in particular, has been shown to exert tremendous influence upon the agenda and content of both other newspapers and television news (Reese & Danielian 1989). Even non-mainstream online news websites have been known to build the agenda of traditional mainstream news outlets (Song 2007). Thus, the PRC can use its own organic news media to engage in inter-media agenda building within the U.S. This would be bolstered by the fact that Western journalists embrace the norm of relying more upon official sources of information as a professional practice (Kiousis *et al*. 2016).

Three Levels of Agenda Building

There are three levels of agenda building. The first level of agenda building is about the transfer of salience for objects like people, nations, and organizations. For instance, during a primary election, a candidate's political campaign will provide journalists with information subsidies to get its candidate on the news agenda. When potential voters hear about the candidate, salience is transferred if they then think about the candidate (Kiousis, Strömbäck & McDevitt 2015). Thus, one might expect a Chinese news outlet to attempt to set the first-level agenda about issues of importance to the state, like the trade war with the U.S.

One compelling study found that Chinese state-sponsored media very successfully engaged in intermedia first-level agenda building about the 2014 Hong Kong Protests. There was a strong positive correlation between the objects discussed by the state media and foreign press coverage of the event (Zhang, *et al*. 2018). Another study of the PRC's *Global Times* showed that it had a first-level, inter-media agenda-building effect upon other global news outlets, including the *Washington Post* (Cui & Wu 2017). Thus, it is reasonable to conclude that the PRC understands agenda building and possesses some skill at influencing the global news agenda.

Second-level agenda building involves the specific attributes of the objects in the first level. In essence, this level of agenda building tells people how to think about the attribute. In the second-level agenda, the builder makes deliberate choices about what attributes to stress and what to ignore. Therefore, the agenda builders can present a very incomplete and biased conceptualization of the object (Kiousis, Popescu & Mitrook 2007). Two types of second-level agenda building have been identified. The first one consists of substantive attributes or traits of people or issues. The second has to do with the affective tone of these attributes, such as whether they are being portrayed as positive or negative (Kihan & McCombs 2007). For instance, when the Bush Administration was building the agenda to justify the invasion of Iraq, it used weapons of mass destruction as a key object. An important attribute of WMD's was the regime's lack of compliance with UN inspections, which the administration used as proof of guilt. What was omitted were alternate explanations of why the regime might have failed to comply with UN inspections (Fhamy*, Wanta, Johnson & Zhang*. 2011). In a study of *Xinhua*, Xi Jingping was framed very positively; however, this only transferred towards a more neutral affective attribute in the U.S. media (Cheng, Golan & Kiousis 2016).

The third level of agenda building is represented by the network connections among elements on an agenda. Essentially the co-occurrence of certain object/attribute combinations with others makes these objects more salient together. For instance, the Bush Administration successfully linked the need for the invasion of Iraq and public fear of weapons of mass destruction. These two objects became so co-salient that they were seen as being integrally tied. This was evidenced in the media, which transferred its salience in the public discourse (Tedesco 2005). In another example, the Don't Let Florida Go to Pot initiative was able to link the fear of teenager marijuana consumption with the use of loopholes present in an amendment that would have legalized the use of medicinal marijuana (Schweickart *et al*. 2016).

Research Questions

Normally agenda-building studies are examined through quantitative content analyses, searching for a correlation between the information subsidy and news coverage about it. The coding categories are decided deductively and then sought out in the material. However, this study is more interested in uncovering what agenda the PRC is trying to build in the U.S. media. This approach would merit an inductive and qualitative approach. Rather than presuming a priori what the coding categories should be, this approach will allow the categories to emerge from the data. Thus, this paper also adds to theoretical development of agenda-building theory by offering the first attempt to examine it qualitatively. Given this, the following research questions are offered.

> RQ1: What agenda issues/objects are most salient in the *Global Times* English language news coverage?
>
> RQ2: What are the substantive attributes that are used to describe each issue/object?
>
> RQ3: Are there any network connections between issues/objects evident?

Methods

News source and sample

The news source chosen was the *Global Times*, an international English language version of the official Chinese Communist Party newspaper the *People's Daily*. It is believed to have a circulation in the tens of millions. Research has shown that it is intensely nationalistic and is uniquely representative of the national leadership's official position (Zeng & Sparks 2020). Only articles with an identified author were accepted for analysis, and articles from Chinese wire services and *Xinhua* were excluded. The unit of analysis was the individual news story, which was analysed line by line. A total sample of (*N*=55) *Global Times* articles were selected from a time frame of 15

April to 25 May 2020. Initially, a sample of 150 articles was randomly chosen; and from these, 75 were randomly chosen for analysis.

Ethnographic Content Analysis

Agenda-building studies are traditionally conducted as quantitative correlational analysis studies. However, this project aims to examine what type of agenda the *Global Times* is trying to build in an inductive manner that searches for the meanings that emerge from the texts. An ideal method for examining propaganda in the form of news texts is Ethnographic Content Analysis (ECA), which is an officially accepted form of qualitative content analysis of texts (Altheide & Johnson 2011). ECA is used to locate, identify, and thematically analyse texts by using ethnographic tools modified for textual analysis; it provides a deeper contextual understanding of meanings that identify theoretical relationships by conceptualizing document analysis as fieldwork (Althiede & Schneider 2013).

The first step in the ECA process is to review the documents and become familiar with their content. The goal is to identify what data researchers can reasonably gain through more rigorous analysis. In the next step, researchers deductively develop a protocol, or a list of questions, items, categories, and variables. A single protocol attempts to capture definitions, meanings, and processes that are important to answer the research questions. Once protocols are defined, the text is analysed for elements that meet the criteria of a specific protocol. These samples are then recorded on a protocols document (Altheide & Schneider 2013) The next step in ECA is data analysis in the form of constant comparison of items discovered for each protocol. A researcher might then examine how often a concept is used and the various contexts in which it is employed. One looks for key differences and extremes between items in the protocols. Analysed data are compiled into summaries (Altheide & Schneider 2013).

For instance, the protocol items reviewed were developed to answer the research questions. Since this study is focused on what the issues objects are, all key people, organizations, and nation states were identified as a protocol item. For substantive attributes, descriptors of these objects, such as quotes from them and adjectives to describe them, were then distinguishable protocol items. Additionally, accompanying pictures, metaphors, and titles of the articles were also listed as protocol items. A protocol item needed to be present in at least five articles to meet the cut-off to be included in further analysis.

Findings

The first research question asked what agenda issues/objects are most salient in the *Global Times* English language news coverage. By far, U.S. President Donald Trump is the most salient object across half of all analysed articles. These included articles that had ostensibly nothing to do with him. Even the U.S. government and the Republican party failed to be mentioned more than five times. The Democratic Party and Democratic politicians are not mentioned at all. The U.S. military was specifically mentioned in seven articles, but all of these were self-classified as military articles. It appears that the Trump Administration is either synonymous with the U.S. government, or it represents an effort to single him out as the locus of all harms to be described later.

The second most salient object was COVID-19, specifically American and allied nations attempts to blame China for malfeasance about its role in the spread of the virus. The third object concept of the U.S. decoupling itself economically from China was salient, but it was often in the context of the COVID-19 conflict. The fourth most salient object is the nation state of China. Interestingly, Xi Jingping is not mentioned at all in any of the articles, and neither are the leaders of China. The nation of China is presented as a monolithic and unified entity. Only in a few places are PRC officials specifically named. In many places where one would expect quotations from high-level government leaders, the *Global Times* often refers to anonymous experts. This was especially true when referencing Chinese military competition with the US. For instance, the title of one article was, "China conducted nuclear tests a U.S. trick to push West-led treaty: Chinese expert" (Lingzhi & Xuanzun 2020).

Research question two asked what the substantive attributes are that are used to describe each issue/object. The substantive attributes associated with Donald Trump are blaming and provoking China. Specifically, he is using COVID-19 to divert attention from his leadership failure on the issue. One article ascribed Trump's actions as a purely political tactic to help him win re-election (Qingqing & Yusha 2020). Another article (Qi 2020) accused Trump 'faking calmness' during the pandemic as a contrivance to bolster faith in his failing political leadership. The editor of the *Global Times* wrote an article describing how Trump's organization of rage against China was diverting attention from his own failures to respond to COVID-19 (Xijin 2020a). In terms of affect, these articles were overtly hostile to Trump personally in a manner that clearly violates the professional norms of Western journalism. One article outright accused Trump of trying to deliberately kill U.S. citizens during the COVID-19 pandemic (Sheng & Qingqing 2020).

The U.S. military is also associated with the attribute of provocateur, purposely agitating a confrontation with China. One article by Xuanzun (2020a) cited how the PRC Navy was prepared for renewed provocations by the U.S. Navy in the South China sea once COVID-19 pandemic was over. It describes in detail how certain U.S. Navy ships engaged in aggressive behaviour with the goal of provoking the PRC. Another article called the U.S. military 'warmongers' for its efforts to bolster Taiwan's defence. It called for the PRC to dramatically expand its nuclear weapons arsenal in response (Xuanzun 2020c).

Attributions about the COVID-19 pandemic were offensively focused on the US. The PRC neither acknowledged nor accepted any responsibility for the pandemic or its own mistakes in responding. An article by Tian (2020) aggressively attacks the American system of 'Federalism' as a reason for the ineffective American COVID-19 response. It discussed how federalism allowed a lack of accountability that does not exist in a superior centralized Chinese system. Even an article discussing an apolitical COVID-19 study from France was turned into an opportunity for blaming, without evidence, American citizens traveling from China as the source of U.S. infections (Keyue 2020). They accused the U.S. of outright theft of medical equipment from other countries, which the PRC then graciously provided (Sheng & Qingqing 2020). In terms of affect, any mention of PRC responsibility was met with a strongly negative emotional reaction. It accused the U.S. of engaging in McCarthyism red-baiting by daring to question China's role in the virus (Yunming & Wenting 2020).

The economic issue of decoupling was clearly salient, especially towards the later end of the sample period. Decoupling is portrayed as unwise and, in some cases, dangerous. It would be more dangerous for the U.S. than for China (Yin 2020). One article threatens the U.S. with reduced tourism revenue from Chinese tourists who spend liberally in the U.S. for decoupling (Feng 2020). Australia is even directly threatened should decoupling intensify. It suggested that remaining strongly aligned with the U.S. would be 'fatal' for Australia-China trade (Xijin 2020b). While decoupling was not desirable, the PRC is signalling its preparedness to thrive despite it. Chinese companies will quickly find ways to overcome hardships and surpass even the US. In the end the U.S. would only 'shoot itself in the foot' (Daye 2020).

Attributions made about the PRC itself take the form of a heroic victim facing unrelenting and needlessly aggressive hostility by the U.S. and its allies. The issue of renewed tariffs against China, which is steadfastly adhering to its new trade agreement, would be unfair and unjust. Despite this, the tariffs would be more harmful to the U.S. than to China (Hongpei & Weiduo 2020). The *Global Times* confidently asserts that U.S. attacks on China are reversing any soft-power gains it had made within China. Chinese students would no longer want to attend U.S. universities and no longer admire the Western way of life (Weiwei 2020). They also proclaimed that China is prepared to become the world leader in global governance, which the United States is abandoning through its actions. China's vastly superior response in dealing with COVID-19, compared to Western nations, was offered as proof (Qingqing & Juecheng 2020).

Attributions of the PRC military were excessively confident to the point of bragging. One report described how the PRC Marines were growing in their capability to project power globally, rivalling the U.S. Marine Corps (Xuanzun 2020b). The growing strength of China at the expense of the west is inevitable and something to be proud of. The *Global Times* interviewed a Harvard political science professor, Arne Westad, who echoed this belief. Westad mentioned that the United States' global leadership was greatly diminished. While he did not believe the pandemic would lead to a cold war, this transfer of power from West to East was essentially inevitable (Yunyi 2020).

Research question three asked if there are any network connections between issues/objects evident. The most obvious network connection is President Trump and the objects of provoking China and decoupling the U.S. and Chinese economies. Both of these issues are strongly salient for the *Global Times* and infused with strongly negative emotions. According to Feng (2020), 'Trump's agenda is to blame China whenever he can". China bashing is necessary for Trump to win re-election, since he botched up during his leadership of COVID (Hongpei & Weidou 2020). Xin (2020) accused Trump of engaging in hybrid warfare against the U.S. with smears about its malfeasance in COVID-19. He ascribed this as a desperate strategy and advocated for the PRC to take a tougher stance against the U.S.

President Trump is also clearly associated with the concept of economic decoupling, which the *Global Times* seems to be very concerned about given its salience. Trump and his subordinates are pushing for decoupling as a scapegoat for America's domestic problems. The Trump administration is portrayed as desiring a new cold war with China, which is unnecessary (Ruohan 2020). An article by a PRC foreign policy expert makes it clear that globalization and remaining coupled to the U.S. are in China's best interests. He admits decoupling is harmful for both sides and should be avoided (Daming 2020). Ironically, the PRC is threatening to economically decouple itself from Taiwan should President Tsai, an advocate of Taiwanese secession, be re-elected (Xijin 2020c).

Discussion

If the goal of the PRC's Three Warfares Strategy is to subvert Western media in order to gain popular support for its foreign policy, it can be inferred from this study that the *Global Times* is building an assertive agenda and promoting it aggressively. It appears to be beginning a strategy of attacking the west and Western values directly. One article by Yiwu (2020), attacks the Western news media directly: "But as these media outlets usually take sides with Western liberal values, some people feel setbacks as their opinions are downplayed by the news organizations" (Yiwu 2020). This could represent the initiation of a strategy advocating for PRC-style governments directly in Western nations.

The *Global Times*' willingness to aggressively attack President Trump on a personal level with emotion is certainly an interesting strategy for an English language newspaper that represents the governing elites of China. The PRC still needs to maintain a working relationship with the U.S. Attacking Trump, rather than just his policies, so directly represents a bold and assertive stance. It could signal several key points. For instance, Trump's hostility towards the American news media is well established, and the PRC could be crafting articles hostile to Trump to increase the chances that it is adopted as an information subsidy by Western journalists. It could also represent the manifestation of a significant threat that Trump does present to the regime. Logically, it probably represents a combination of both.

Based upon the evidence presented here, the PRC appears to have little tolerance for any criticism about its role in COVID-19. This reaction is not reasonable since the PRC's delays in implementing the quarantine and failure to communicate with the world could be ascribed to confusion about a new virus and internal miscommunication. The overwhelming consensus in the crisis communication literature, which PRC communications experts understand, indicate that a simple acknowledgement and apology would have sufficed to resolve the issue. Their extreme sensitivity about this issue signifies that something more significant is at play here. In the context of the Three Warfares Strategy, refusal to accept criticism over COVID-19 could represent an imperative to portray the PRC's political system as superior to those of the West.

The issue of economic decoupling with the U.S. and Australia is also a salient issue on the PRC agenda. It clearly does not desire to engage in decoupling but is strongly signalling that it has plans to manage the fallout from decoupling. One article stated about Australia, "If Australia blindly limits China-Australia scientific research cooperation, it will further hurt its own economic interests" (Lei 2020, p. 1) The tone on the issue of decoupling is strongly defiant: the West needs the PRC more than the PRC needs the West. From a Three Warfares perspective, this could lead to future propaganda themes blaming the West for any future economic recession or depression because it foolishly decoupled from the PRC.

The *Global Times* agenda indicates a pre-occupation with provocation by the U.S. and its allies. They levelled direct threats against Australia, Taiwan, and Hong Kong secessionists for engaging in provocative behaviour. Any action by the U.S. military that was covered in the *Global Times* was framed as intentional provocation. Part of the reason for the fury at President Trump is that the PRC perceives his behaviour as strongly provocative. Within U.S. political culture, competitors routinely engage in provocative behaviour to the point that it goes unnoticed by the public. It is very probable that intolerance for provocation is rooted in some aspect of Chinese culture.

Finally, the *Global Times*' agenda is offensive in nature. It vociferously attacks any party that is at odds with PRC interests. Coverage of its own military is overly optimistic about its capabilities and is slightly menacing towards Taiwan and the U.S. Several articles directly attack the legitimacy of the Western democratic model of government. The articles overtly tout the superiority of the Chinese government's centralized control during the COVID-19 pandemic. They attack the legitimacy of the American electoral process in several articles. One article directly blames elite U.S. politicians for wilfully allowing Americans to die from COVID-19 for short-term political gamesmanship (Rangu 2020). The authors seem to imply that if the U.S. had a Chinese style of government it would be seriously improved.

Contributions

This article contributes to the literature in both a theoretical and practical sense. It represents the first attempt at analysing agenda-building theory qualitatively. Qualitative research can improve the theory by adding a method of exploring the meanings created by an agenda in depth not afforded by quantitative analysis. Secondly, it adds an initial and exploratory study of the PRC's Three Warfare Strategy; which is currently discussed in only two articles despite its obvious importance. Establishing that the PRC is on the offensive in its information warfare activities provides a starting point for future research.

Conclusion

The agenda being built by the *Global Times* provides some insights into the political mindset and goals of the PRC. How effective its agenda building is currently would be difficult to ascertain given the current administration's hostility with the press. It would be impossible to determine if negative news coverage about the administration was actually caused by the *Global Times* or a Chinese news site. The current administration's strategy of blaming China for COVID-19 as well as threatening economic decoupling is achieving a profoundly negative reaction from the PRC.

It is highly probably that as tensions between the U.S. and the PRC become increasingly tense, the news agenda will grow even more offensive. It is probable that the PRC will attempt to challenge the legitimacy and efficacy of the Western democratic system directly within those countries. The PRC would probably couple this with information warfare to promote its political model directly to American citizens. Additionally, the PRC's inability to deal with any criticism of its behaviour represents a profound weakness in its strategy. Governments make mistakes, and it is often best to acknowledge them, apologize, and move on. However, evidence from the *Global Times* indicates that the PRC media machine will do all in its power to protect the ruling elite, even when it is disadvantageous to do so.

References

Altheide, DL & Johnson, JM 2011, 'Reflections on interpretive adequacy in qualitative research', *The Sage handbook of qualitative research*, eds. NK Denzin, & YS Lincoln, Sage, Thousand Oaks, CA, US, pp. 581–610.

Altheide, DL & Schneider, CJ 2013, '*Qualitative media analysis',* Sage, Thousand Oaks, CA, US.

Bisht, NSP, Jain, R & Gambhir, V 2019, 'Doklam Plateau and Three Warfares Strategy', *China Report*, vol. *55, no.* 4, pp. 293.

Burnett, NF 1989, ‘Ideology and propaganda: Toward an integrative approach’, *Propaganda: A Pluralistic Perspective*, ed. T Smith, Praeger, New York, NY, US, pp. 2-6.

Chang, J and Ren, H 2018, ‘The powerful image and the imagination of power: The “new visual turn” of the CPC’s propaganda strategy since its 18th National Congress in 2012’, *Asian Journal of Communication*, vol. 28, no. 1, pp. 1–19, doi: 10.1080/01292986.2017.1320682.

Cheng, Z, Golan, GJ & Kiousis, S 2016, ‘The Second-Level Agenda-Building Function of the Xinhua News Agency’, *Journalism Practice*, vol. 10, no. 6, pp. 744–762, viewed 16 October 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=ufh&AN=118888403&site=eds-live&scope=site>.

Cui, D & Wu, F 2017, ‘Inter-media agenda setting in global news production: Examining agenda attributes in newspaper coverage of the MH370 incident in the U.S., China, and Hong Kong’, *Asian Journal of Communication*, vol. 27, no. 6, pp. 582–600, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=ufh&AN=125437673&site=eds-live&scope=site>.

Daming, D 2020, ‘Will China-US decoupling continue after COVID-19?’ *Global Times,* viewed 8 June 2020, <http://www.globaltimes.cn/content/1188053.shtml>.

Daye, C. 2020, ‘Facing US tech war, China needs to turn to homegrown innovations.’ *Global Times*, viewed 8 June 2020 <http://www.globaltimes.cn/content/1189347.shtml>.

Denham, B 2010, ‘Toward conceptual consistency in studies of agenda-building processes: A scholarly review’, *Review of Communication*, vol. 10, no. 4, pp. 306–23, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=ufh&AN=66418311&site=eds-live&scope=site>.

Iasiello E 2016, ‘China’s Three Warfares Strategy mitigates fallout from cyber espionage activities’, *Journal of Strategic Security*, vol. 9, no. 2, p. 45, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=edsjsr&AN=edsjsr.26466776&site=eds-live&scope=site>.

Fahmy, SS, Wanta, W, Johnson, TJ & Zhang, J 2013, ‘The path to war: Exploring a second-level agenda-building analysis examining the relationship among the media, the public and the president’, *International Communication Gazette*, vol. 73, no. 4, pp. 322–42, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=edselc&
AN=edselc.2-52.0-79958699595&site=eds-live&scope=site>.

Feng, D 2020, ‘China-US relations at a low ebb but decoupling not the answer’, *Global Times,* viewed 8 June 2020, <https://www.globaltimes.cn/content/1187244.shtml>.

Hongpei, Z & Weiduo, S 2020, ‘Fresh tariff threat adds woes to US economy’, *Global Times*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1187726.shtml>.

Hughes, A & Dann, S 2009, ‘Political marketing and stakeholder engagement’, *Marketing Theory*, vol. 9, no. 2, pp. 243–56, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=buh&AN=4134683&site=eds-live&scope=site>.

Jowett, GS & O’Donnell, V 2015, *Propaganda and persuasion*, Sage, Los Angeles, CA, US.

Keyue, X, 2020, ‘Apolitical scientific tracing of virus urged after earlier cases found in France’, *Global Times,* viewed 8 June 2020, <https://www.globaltimes.cn/content/1187460.shtml>.

Kihan, K & McCombs, M 2007, ‘News story descriptions and the public’s opinions of political candidates’, *Journalism & Mass Communication Quarterly*, vol. 84, no. 2, pp. 299–314, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=eft&AN=507988412&site=eds-live&scope=site>.

Kiousis, S, Popescu, C & Mitrook, M 2007, ‘Understanding influence on corporate reputation: An examination of public relations efforts, media coverage, public opinion, and financial performance from an agenda-building and agenda-setting perspective’, *Journal of Public Relations Research*, vol. 19, no. 2, pp. 147–65, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=buh&AN=25075121&site=eds-live&scope=site>.

Kiousis, S, Ragas, MW, Kim, JY, Schweickart, T, Neil, J & Kochhar, S 2016, ‘Presidential agenda building and policymaking: Examining linkages across three levels’, *International Journal of Strategic Communication*, vol. 10, no. 1, pp. 1–17, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=ufh&AN=112998563&site=eds-live&scope=site>.

Kiousis, S, Strömbäck, J & McDevitt, M 2015, ‘Influence of issue decision salience on vote choice: Linking agenda setting, priming, and issue ownership’, *International Journal of Communication (19328036)*, vol. 9, pp. 3347–68, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=ufh&AN=110802624&site=eds-live&scope=site>.

Lariscy, R, Avery, E & Sohn, Y 2010, ‘Health journalists and three levels of public information: issue and agenda disparities?’, *Journal of Public Relations Research*, vol. 22, no. 2, pp. 113–35, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=buh&AN=49147025&site=eds-live&scope=site>.

Lauk, EPP & Harro-Loit, H 2017, ‘Journalistic autonomy as a professional value and element of journalism culture: The European perspective’, *International Journal of Communication (19328036)*, vol. 11, pp. 1956–74, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=ufh&AN=126812985&site=eds-live&scope=site>.

Lee, SY & Riffe, D 2017, ‘Who sets the corporate social responsibility agenda in the news media? Unveiling the agenda-building process of corporations and a monitoring group’, *Public Relations Review*, vol. 43, no. 2, pp. 293–305, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=edselp&AN=S0363811115301855&site=eds -live&scope=site>.

Lei, Y 2020, ‘Limited scientific cooperation with China would hurt Australia’, *Global Times*. viewed 8 June 2020, <http://www.globaltimes.cn/content/1195484.shtml>.

Lin, C 2017, ‘From “poison” to “seeder”: the gap between propaganda and *xuanchuan* is cultural’, *Asian Journal of Communication*, vol. 27, no. 5, pp. 451–63, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=ufh&AN=124333224&site=eds-live&scope=site>.

Lingzhi, F & Xuanzun, L 2020, “‘China conducted nuclear tests” a US trick to push West-led treaty: Chinese expert’, *Global Times. v*iewed 8 June 2020, <https://www.globaltimes.cn/content/1185857.shtml>.

Livermore, D 2018, ‘China’s “Three Warfares” in theory and practice in the South China Sea’, *Georgetown Security Studies Review.* viewed 8 June 2020, <https://georgetownsecuritystudiesreview.org/2018/03/25/chinas-three-warfares-in-theory-and-practice-in-the-south-china-sea/>.

McCombs, M 2014, *Setting the agenda: The mass media and public opinion*, Polity, Cambridge, UK.

Pei-Ling, L 2014, ‘The impact of Chinese military philosophy on the development of propaganda tactics’, *China Media Research*, vol. 10, no. 4, pp. 39–47, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=ufh&AN=99659779&site=eds-live&scope=site>.

Qi, W 2020, ‘White House dishonesty, “fake calmness” to spark political uproar’, *Global Times*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1188047.shtml>.

Qingqing, C & Juecheng, Z 2020, ‘Global test of governance is on wrong time for West to reopen society, may hurt China: expert’, *Global Times*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1188064.shtml>.

Qingqing, C & Yusha, Z 2020, ‘US blame game plotted for US election’, *Global Times,* viewed 8 June 2020, <https://www.globaltimes.cn/content/1187482.shtml>.
Rangu, F 2020, ‘US elites try to set up China as pandemic scapegoat with unauthorized material: expert.; *Global Times*, viewed 8 June 2020, <http://www.globaltimes.cn/content/1188312.shtml>.

Reese, SD & Danielian, LH 1989, ‘Intermedia influence and the drug issue: Converging on cocaine’, *Communication campaigns about drugs,* ed. PJ Shoemaker, Lawrence Erlbaum, Hillsdale, NJ, US, pp. 47-56.

Ruohan, L 2020, ‘China-US decoupling “impossible, impractical”’, *Global Times,* viewed 8 June 2020, <http://www.globaltimes.cn/content/1146595.shtml>.

Schaeffer, K 2020, ‘Nearly three-in-ten Americans believe COVID-19 was made in a lab’, *Pew Research Center*. viewed 8 June 2020, <https://www.pewresearch.org/fact-tank/2020/04/08/nearly-three-in-ten-americans-believe-covid-19-was-made-in-a-lab/>.

Schweickart, T, Neil, J, Kim, JY & Kiousis, S 2016, ‘Time-lag analysis of the agenda-building process between White House public relations and congressional policymaking activity’, *Journal of Communication Management*, vol. 20, no. 4, pp. 363–80, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=edsemr&AN=edsemr.10.1108.JCOM.01.2016.0001&site=eds-live&scope=site>.

Schweickart, N, Zhang, T, Lukito, J, Kim, JY, Golan, G & Kiousis, S 2018, ‘The dash FOR GAS: Examining third-level agenda-building and fracking in the United Kingdom’, *Journalism Studies*, vol. 19, no. 2, pp. 182–208, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=ufh&AN=127339520&site=eds-live&scope=site>.

Sheng, Y, & Qingqing, C 2020, ‘Washington ‘launches COVID-19 war against its own people’, *Global Times*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1186818.shtml>.

Song, Y 2007, ‘Internet news media and issue development: A case study on the roles of independent online news services as agenda-builders for anti-US protests in South Korea’, *New Media & Society*, vol. 9, no. 1, pp. 71–92, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=psyh&AN=2007-01484-002&site=eds-live&scope=site>.

Tedesco, JC 2005, ‘Issue and strategy agenda setting in the 2004 presidential election: Exploring the candidate–journalist relationship’, *Journalism Studies*, vol. 6, no. 2, pp. 187–201, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=ufh&AN=16970522&site=eds-live&scope=site>.

Tian, S 2020, ‘Nobody accountable in US despite rising virus deaths’, *Global Times,* viewed 8 June 2020, <https://www.globaltimes.cn/content/1185859.shtml>.

Wang, H, Sparks, C & Yu, H 2018, ‘Popular journalism in China: A study of China youth Daily’, *Journalism: Theory, practice, and criticism*, vol. 19, no. 9–10, pp. 1203–19, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=mzh&AN=2019870381&site=eds-live&scope=site>.

Weiwei, Z 2020, ‘Blame game diminishes Western appeal to Chinese people’, *Global Times*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1187809.shtml>.

Wilbur, D 2017, ‘Propaganda’s place in strategic communication: The case of ISIL’s Dabiq Magazine’, *International Journal of Strategic Communication*, vol. 11, no. 3, pp. 209–23, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=edselc&AN=edselc.2-52.0-85019090833&site=eds-live&scope=site>.

Xin, L 2020, ‘China needs tougher response to US smears: experts’ *Global Times*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1188409.shtml>.

Xijin, H 2020a, ‘Trump should be political Batman for US in crisis’, *Global Times’*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1187210.shtml>

——2020b, 'Australia should distance itself from a possible new China-US "cold war"', *Global Times*, viewed 8 June 2020, <http://www.globaltimes.cn/content/1189356.shtml>.

——2020c, 'Cross-Straits ties likely to see "a decoupling future" as Tsai enters 2nd term', *Global Times*, viewed 8 June 2020, <http://www.globaltimes.cn/content/1188936.shtml>.

Xuanzun, L 2020a, 'China conducts naval drills in S.China Sea, prepares for post-pandemic US military provocations', *Global Times*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1187466.shtml>.

——2020b, 'China deploys AEW, anti-submarine aircraft on South China Sea's *Yongshu* Reef: Report', *Global Times*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1188427.shtml>.

——2020c, 'China urged to expand nuclear arsenal to deter US warmongers', *Global Times*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1187775.shtml>.

Yin, C 2020, 'Neither de-Sinicization nor de-Americanization possible after pandemic', *Global Times,* viewed 8 June 2020, <http://www.globaltimes.cn/content/1188869.shtml>.

Yiwu, Z 2020, 'Why both police and protesters attack Western media?', *Global Times, v*iewed 8 June 2020, <http://www.globaltimes.cn/content/1191364.shtml>.

Yuan, Li 2020, 'With selective Coronavirus coverage, China builds a culture of hate', *The New York Times*, viewed 8 June 2020, <https://www.nytimes.com/2020/04/22/business/china-coronavirus-propaganda.html>.

Yunming, Y & Wenting, X 2020, 'Post-pandemic animosity by US', *Global Times*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1187803.shtml>.

Yunyi, B 2020, 'China-US Cold War 2.0 unlikely after pandemic ends', *Global Times*, viewed 8 June 2020, <https://www.globaltimes.cn/content/1187744.shtml>.

Zeng, W & Sparks, C 2011, 'Popular nationalism: Global times and the US–China trade war', *International Communication Gazette*, vol. 82, no. 1, pp. 26–41, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=edselc&AN=edselc.2-52.0-85074604998&site=eds-live&scope=site>.

Zhang, T, Khalitova, L, Myslik, B, Mohr, TL, Kim, JY & Kiousis, S 2018, 'Comparing Chinese state-sponsored media's agenda-building influence on Taiwan and Singapore media during the 2014 Hong Kong Protest', *Chinese Journal of Communication*, vol. 11, no. 1, pp. 66–87, viewed 5 August 2020, <http://search.ebscohost.com.proxy.mul.missouri.edu/login.aspx?direct=true&db=ufh&AN=128484710&site=eds-live&scope=site>.

Offensive Cyberspace Operations and Zero-days: Anticipatory Ethics and Policy Implications for Vulnerability Disclosure

G Huskaj[1,2], RL Wilson[3,4]

[1]Department of Military Studies
Swedish Defence University, Stockholm, Sweden

[2]School of Informatics University of Skövde, Skövde, Sweden

[3]Department of Philosophy and Computer and Information Sciences
Towson University, Towson, Maryland, United States

[4]Hoffberger Center for Professional Ethics
University of Baltimore, Baltimore, Maryland, United States

E-mail: gazmend.huskaj@fhs.se; wilson@towson.edu

Abstract: *This article addresses the question under which circumstances zero-day vulnerabilities should be disclosed or used for offensive cyberspace operations. Vulnerabilities exist in hardware and software and can be seen as a consequence of programming errors or design flaws. The most highly sought are so-called zero-day-vulnerabilities. These vulnerabilities exist but are unknown and, when exploited, enable one way of entry into a system that is otherwise not thought possible. Therefore, from an anticipatory ethics perspective, it is important to understand in what cases zero-days should be disclosed or not.*

Keywords: *Anticipatory Ethics, Information Systems, Offensive Cyberspace Operations, Stuxnet, Vulnerabilities, Zero-Days*

Introduction

This article addresses the following question: under which circumstances should zero-day vulnerabilities be disclosed or used for offensive cyberspace operations? Vulnerabilities exist in hardware and software and can be seen as a consequence of programming errors or design flaws. One example is the 1988 Morris worm. It "attacked vulnerable services including fingerd (used to find information about computer users) and sendmail (used to send email) [and] when it attacked fingerd, it sent a 536-byte request to C code using a vulnerability that provided a buffer with only 512 bytes of space; the resulting overflow allowed the worm's code to execute on the target" (O'Leary 2019, p. 51). This is what is also known as a 'buffer overflow', an offensive method. Threat actors can exploit these vulnerabilities for offensive cyberspace operations to gain otherwise unintended access to information systems, resources, and/or stored information (Huskaj 2019). In other words, they can be used to impact the confidentiality, integrity, and availability in information systems to generate deny, disrupt, degrade, destroy, or manipulate effects. The reasons for exploiting these vulnerabilities may vary: to steal intellectual property or to plant 'logic-bombs' in critical infrastructure, to employ for intelligence purposes, or to use for offensive cyberspace

ISSN 1445-3312 Print/ISSN 1445-3347 Online

operations. The most highly sought-after vulnerabilities are so-called 'zero-day' vulnerabilities. These vulnerabilities exist but are unknown; and when exploited, they enable one way of entry into a system that is otherwise not thought possible. This is also why zero-day vulnerabilities are very popular among criminal organizations, states, and state-sponsored advanced persistent threats.

The article is structured as follows. First, discussions about zero-days in the scientific community are presented and followed by examination of vulnerabilities and vulnerability disclosure. Next, the cases are presented. Finally, understanding ethics and related ethical discussions are considered, followed by policy-related ethical analysis and the conclusions of anticipatory ethical recommendations/policy.

Discussions about Zero-Days in the Scientific Community

Surveying the literature on zero-days in hardware and software is important to generate insights about the focus of that research. Scopus® was queried using the search terms (”zero day*” OR “zero-day”) AND (hardware OR software). The database returned 304 document results. Abstracts without author names were removed, including doubles, and the conference version of this article was also removed. The 291 remaining abstracts were manually reviewed to identify any possible noise in the data set. The review revealed two articles that were beyond the scope. The remaining 289 abstracts were analysed employing the Computational Literature Review (CLR). The CLR was developed by Mortenson & Vidgen (2016) to automate the analysis of research articles by considering the analysis of impact through citations; structure by co-authorship networks; and content through topic-modelling of the abstracts. However, for this research, only topic modelling is considered. The results of topic modelling on the 289 remaining abstracts are depicted in **Table 1**.

Topic No.	Insights from the topic	Clustered into
1	Software vulnerabilities, bugs, how attackers exploit them, and related security	Vulnerabilities
2	Attacks exploiting zero-days on data networks	Attacks
3	Detecting malware that exploit zero-days	Detection
4	Zero-day attacks exploiting vulnerabilities	Attacks
5	Security vulnerabilities in software	Vulnerabilities
6	Identifying software vulnerability for cybersecurity	Identify
7	Public-private partnerships, zero-days, and cybersecurity to mitigate attacks	Public-private partnerships
8	Attack detection in networks	Detection
9	Searching for zero-days through techniques such as fuzzing	Identify
10	Software vulnerability threats to security	Threats
11	Zero-day attacks and security	Attacks
12	Zero-day vulnerabilities/attacks in various systems	Vulnerabilities and attacks

Table 1: Topic modelling of 289 abstracts

Table 1 reveals that zero-days are discussed from multiple perspectives: Topics 2, 4, and 11 are clustered into attacks. In this cluster, topic 2 discusses attacks that exploit zero-days, while topic 4 has the perspective of zero-day as an attack that exploits a vulnerability. Topic 11, like topic 4, considers a zero-day as an attack, but it also discusses security. Topics 3 and 8 are about detection: detecting malware that exploits zero-days (topic 3) and attacks in networks (topic 8). Topics 6 and 9 are about identifying software vulnerabilities for cybersecurity (topic 6) and using various techniques, such as fuzzing, to search for zero-day vulnerabilities in software (topic 9). Topic 7 discusses public-private partnerships in relation to zero-days and cybersecurity to mitigate attacks. Topic 10 considers zero-days as software vulnerability threats to security. Finally, topics 1 and 5 are about software vulnerabilities and security vulnerabilities in software.

One direct conclusion from the review is that zero-days are discussed in the context of software and not hardware. This is reasonable because software manages and directs the operation of hardware. Another conclusion is that research about public-private partnerships and zero-days is sparse. Finally, there is lack of research on the ethics of vulnerability disclosure.

No. Cites	Authors	Years	Title and description	Topic
413	Grace *et al*.	2012	RiskRanker: scalable and accurate zero-day android malware detection. This article discusses an automated system to analyse malicious apps. It found 322 zero-days after scanning through 118 318 in various Android markets.	10
374	Cárdenas *et al*.	2011	Attacks against process control systems: risk assessment, detection, and response. This article discusses how to detect computer attacks on SCADA systems by focusing on the objective of the attack. Then, automatic response mechanisms ensure the system is not driven to an unsafe state.	6
361	Singh *et al*.	2004	Automated worm fingerprinting. Unrestricted connectivity and software homogeneity are risks. An automated approach to detect unknown worms and viruses is proposed.	9
190	Seshadri *et al*.	2007	SecVisor: a tiny hypervisor to provide lifetime kernel code integrity for commodity Operating Systems. It is a tiny hypervisor that defends against attackers with knowledge of zero-day kernel exploits.	10
104	Chen *et al*.	2015	Finding unknown malice in 10 seconds: mass vetting for new threats at the Google-play scale. MassVet is a technique for mass vetting apps. The results of vetting nearly 1.2 million apps form 33 app markets revealed over 20 likely zero-day malware.	10
103	Lu *et al*.	2010	BLADE: an attack-agnostic approach for preventing drive-by mal- ware infections. The BLADE system protects against code obfuscation and zero-day threats. The system was evaluated using various versions of Internet Explorer and Firefox and it blocked all of the malware install attempts in 1934 active malicious URLs.	4

Table 2: Abstracts with more than 100 citations.

Another way to generate insights into the related literature is through numbers of citations. There were six abstracts that had more than 100 citations. **Table 2**, above, depicts these. Analysing the number of citations and descriptions reveals that scientists focus on developing automated systems to detect, respond, fingerprint, defend, mass vet, and prevent software with zero-days.

Vulnerabilities and Vulnerability Disclosure

This section describes vulnerabilities and the vulnerability disclosure process (VEP). According to Pfleeger & Pfleeger (2012, p. 10), “A vulnerability is a weakness in the system, for example, in procedures, design, or implementation, that might be exploited to cause loss or harm”. Vulnerabilities enable access to information systems. In the Stuxnet case, four zero-day vulnerabilities were used, and a fifth vulnerability was already known. **Table 3**, below, shows the zero-days and the non-zero-day vulnerability. The offensive methods to generate denial effects are also depicted in **Table 3**.

Name	VUL TYP	C	I	A	Access CXTY	AUTH	Gained Access	CVE
.lnk	EC	CPLT	CPLT	CPLT	M	N/R	None	CVE-2010-2568
Print Spooler Service Impersonation	EC	CPLT	CPLT	CPLT	M	N/R	None	CVE-2010-2729
Win32k Keyboard Layout	GP	CPLT	CPLT	CPLT	L	N/R	None	CVE-2010-2743
Task Scheduler	GP	CPLT	CPLT	CPLT	L	N/R	None	CVE-2010-3338
Server Service	ECO	CPLT	CPLT	CPLT	L	N/R	Admin	CVE-2008-4250

Notes: C = Confidentiality; I = Integrity; A = Availability; AUTH = Authentication; CPLT = Complete; CXTY = Complexity; EC = Execute Code; ECO = Execute Code Overflow; GP = Gain Privileges; L = Low; M = Medium; N/R = Not Required; VUL TYP = Vulnerability Type

Table 3: Zero-days and non-zero-day vulnerabilities

The problem for vendors is whether customers should be notified of the existence of these vulnerabilities. The purpose of the VEP is “to balance equities and make determinations regarding disclosure or restriction when the USG obtains knowledge of newly discovered and not publicly known vulnerabilities in information systems and technologies” (White House 2017, p. 1). Managing vulnerabilities responsibly is important because of the “significant economic, privacy and national security implications [they can have] when exploited” (White House 2017, p. 2). The process is not led by a single agency; rather, it is “coordinated by the National Security Council (NSC) staff so that multiple agency viewpoints can be considered, informed by the full input and consideration of the interagency experts” (White House 2017, p. 2). The process has a threshold for vulnerabilities, whether they are worthy to be considered for the VEP or not. The threshold is that the vulnerability “must be both newly discovered and not publicly known” (White House 2017, p. 5). If the vulnerability meets this requirement, it is then submitted to the VEP Executive Secretariat. The submission “will include, at a minimum, information describing the vulnerability, identification of the vulnerable products or systems, and a recommendation on dissemination of the vulnerability information” (White House 2017, p. 7). Next, equity discussions take place to disseminate or restrict information about the vulnerability. If consensus is reached, the Equities Review Board (ERB) ratifies a recommendation for further actions such as “sharing, restricting or reassessing” (White House 2017, p. 7).

The Cases

Three cases are considered: Iraq, Syria, and Iran. Each case is further described below.

Iraq

In 1981, an Israeli air operation dubbed "Operation Opera" was executed to destroy Saddam Hussein's nuclear plutonium reactor. In 1977, Israel discovered that Iraq was developing a nuclear reactor. At their disposal, the Israeli Air Force had F-4 Phantoms and Skyhawks which "were not capable of flying the over 1,000 miles into enemy territory and returning safely" (TOI STAFF 2019). However, during the 1978-1979 Islamic Revolution in Iran, 75 U.S. F16s intended for the country were cancelled and redirected to Israel (TOI STAFF 2019). Accounts of the type of aircraft used in the operation differed. Shipler (1981) noted that "American military analysts said that the bombing was apparently done by American-made F-4 Phantoms escorted by F-15's." Evans (2017), on the other hand, notes "fourteen American-built F-16 fighter aircraft had taken off from an Israeli airstrip in the Negev". However, according to the IAF commander in 1981, Major General David Ivry, "eight aircraft instead of the originally planned four" were used (TOI STAFF 2019). These were F16s.

Ten Iraqi soldiers and one French technician were killed in the attack (BBC 2006). The Italian Government reported that "none of the 20 Italian technicians at the project had been injured" (Lewis 1981). The consequences of the attack were condemnations from the British Foreign Office, the Secretary General of the United Nations, and the Soviet Union (Lewis 1981); and, according to a U.S. Intelligence assessment, the "attack has hurt US interests" (Evans 2017).

Syria

In 2007, an Israeli air operation dubbed "Operation Outside the Box" destroyed Bashar al-Assad's suspected nuclear reactor. In 2004, Israeli intelligence believed North Korea was helping Syria to build a nuclear reactor (BBC 2018). In 2006, Israel had additional information that confirmed the details of the situation. At 22:30 on 5 September 2007, four F15 and four F16 aircraft flew to Deir al-Zour and bombed the nuclear reactor (BBC 2018). Israel "verified that the reactor was destroyed 'beyond any chance of rehabilitation'" (Opall-Rome 2018) on 6 September 2007 at 02:30. There is no information about the death toll from the attack.

Iran

Since the 1950s, Iran has shown interest in nuclear technology. However, during the Revolution, those plans were scrapped by Ayatollah Ruhollah Khomeini (NTI 2020). Later, in 1984, Iran once again began the development of the Bushehr facility (NTI 2020). U.S. intelligence agencies suspected the Iranians were using the civilian program as a cover to develop nuclear weapons; and "on 30 April 2018, Israeli Prime Minister Benjamin Netanyahu delivered a presentation in which he revealed the seizure of over 100,000 documents by Israeli intelligence from what he called 'Iran's secret atomic archives'" (NTI 2020).

Iran was confronted with these allegations, but the Iranian government stated that the purpose was to use the technology for nuclear power plants. However, the U.S., Israel, and several other countries were not convinced. Allegedly, Israel pushed to intervene and had meetings with the U.S. government, which decided against using force. The preferred course of action was to use cyber capabilities to affect the program.

Preparing for a cyber operation likely required a lot of intelligence collection on the program (including facilities, people, and uranium centrifuges). Once that intelligence was collected, it directed the requirements to create a testbed. It revealed, amongst other things, that the information systems managing the centrifuges were air gapped, in other words, not connected to the Internet. Further intelligence also showed they used the Microsoft Windows operating system (OS) and Simatic WinCC SCADA systems to operate the centrifuges (Naraine 2010b).

Identifying potential vulnerabilities in the OS and the SCADA systems led to the development of a testbed that used the same versions of the OS and the SCADA systems. The development of the software was spread around the research labs of the U.S. government for operation security (OPSEC) reasons. The first step was to solve the problem of the air-gapped infrastructure. The second was to execute the software in the air-gapped infrastructure. The third was to change the rotation frequency of the centrifuges. The fourth was to accomplish all of this covertly.

The air-gapped problem required a human to physically insert a USB-memory stick in an information system operating the centrifuges. The requirement for the software was for it to execute on the target system without human interaction.

Identifying the potential vulnerabilities led to the discovery of at least four zero-day vulnerabilities. Naraine (2010b) stated that "The attackers behind the recent Stuxnet worm attack used four different zero-day security vulnerabilities to burrow into—and to spread around—Microsoft's Windows operating system". The four zero-days exploited a .lnk file vulnerability, "a remote code execution vulnerability as well as two local privilege escalation vulnerabilities" (Murchu 2010).

The .lnk file vulnerability allowed "local users or remote attackers to execute arbitrary code via a crafted (1) .LNK or (2) .PIF shortcut file, which is not properly handled during icon display in Windows Explorer" (NVD 2010a). The issue at hand was improper input validation. Improper input validation occurs "when software does not validate input properly, [and] an attacker is able to craft the input in a form that is not expected by the rest of the application. This will lead to parts of the system receiving unintended input, which may result in altered control flow, arbitrary control of a resource, or arbitrary code execution" (MITRE 2019).

The remote code execution vulnerability "does not properly validate spooler access permissions, which allows remote attackers to create files in a system directory, and consequently execute arbitrary code" (NVD 2010b). The same issue that was identified above was here: improper input validation.

The two local privilege escalation exploits targeted vulnerabilities in Keyboard layout file and Task Scheduler (Naraine 2010a). The first one loaded an index without verification, allowing "the malware to force the system's kernel to execute code controlled from the user area" (Xmco 2011, p. 16). The second one modified the task file to create a CRC32 collision, allowing an attacker to "execute arbitrary commands with SYSTEM privileges" (WebDEViL 2010).

The operation consisted of two phases: one in 2007 and one in 2010. The first attack, in 2007, sought to overpressure centrifuges while the second attack in 2010 involved changing the rotation frequency (Langner 2013). Changing the rotation frequency of the centrifuges required specific knowledge of the SCADA systems controlling the centrifuges. Doing this covertly meant having a high classification on the operation, coupled with high OPSEC. The code of the second attack in 2010 revealed it altered the rotation speed of the enrichment centrifuges, which resulted in Iran's failure to enrich uranium. Furthermore, unconfirmed information states that the code would have never been found if certain actors stressed the development of the 2010 code.

Understanding Ethics

This section describes the ethical issues with the disclosure of zero-day vulnerabilities. Zero-day vulnerabilities present the possibility of developing exploits which can be used to generate, deny, disrupt, degrade, destroy, or manipulate effects. Furrow (2005) identifies the focus of ethical analysis as involving a series of factors. He states that ethics is related to evaluating actions, and actions are performed by those capable of being moral agents. He writes, "When we evaluate an action, we can focus on various dimensions of the action. We can evaluate the person who is acting, the intention or motive of the person acting, the nature of the act itself, or the consequences" (Furrow 2005, p. 44).

Two particular variations are presented here. The first one is the ethical issues related to zero-day vulnerabilities, which are based upon the idea that those who discover them have taken actions to discover them, and these actions are an extension of what a person intends. The second is that the actions to discover zero-day vulnerabilities are capable of being evaluated based upon the intentions and actions of the people engaged in those activities, as well as the outcomes of their actions. Likewise, the action to disclose or not to disclose is capable of being evaluated based on those responsible for making that decision. Applying Furrow's (2005) distinctions to zero-day vulnerabilities leads researchers to three possible levels of ethical evaluation. First are the actions of a person performing actions to discover zero-day vulnerabilities, and the action/decision to disclose it or not? Second are the intentions of a person's actions to discover zero-day vulnerabilities? Third, is the nature of the act or the consequences of the actions intended by the person(s) discovering zero-day vulnerabilities and their action/decision to disclose them or not?

The actions of agents discovering zero-day vulnerabilities are subject to ethical evaluation based on the actions of the person(s) deciding to disclose or not, the intentions or motives of the person(s), and the consequences produced by disclosing or not disclosing zero-day vulnerabilities. Using Furrow's (2005) distinctions, it is ultimately the person or persons who are deciding whether to disclose or not disclose zero-day vulnerabilities that are subject to moral evaluation. Identifying the ethical issues with zero-day vulnerabilities requires asking four questions:

1) What actions are performed when disclosing or not disclosing a zero-day vulnerability?
2) What is the character of the person(s) taking those decisions?
3) What are the intentions of those deciding whether disclosure should or should not occur?
4) What are the consequences of disclosing or not disclosing zero-day vulnerabilities?

The ethical issues with zero-day vulnerabilities are the result of how they are exploited and the potential consequences if they are not disclosed. The purpose of withholding knowledge of a zero-day vulnerability is to achieve tactical, operational, or strategic goals by that person or persons. In this case, there is the controller of a zero-day vulnerability and those affected by the exploit of zero-day vulnerabilities. The intention of the person or persons engaged in exploiting zero-day vulnerabilities involves instrumental reasoning and establishing a purpose for exploiting zero-day vulnerabilities, as well as a goal (such as degrading a nuclear weapon program). Next is the adversary's nuclear weapon program which poses an existential threat affected by the purpose of zero-day vulnerabilities, which involves the technical issue of affecting the information systems responsible for uranium-enrichment centrifuges. The interaction between the technical use of a software exploiting a zero-day vulnerability by the person or persons to affect a uranium enrichment centrifuge and the technical effect of the exploit as it affects information systems and industrial and control systems are where ethical issues, with disclosing or not disclosing zero-day vulnerabilities, arise. A preliminary ethical analysis can be developed by applying standard ethical principles. With appropriate space and time, these standard ethical principles can be applied to the intentions of the person(s) deciding whether to disclose or not disclose zero-day vulnerabilities, to the actions of disclosing or not disclosing zero-day vulnerabilities, and to the outcomes of disclosing or not disclosing zero-day vulnerabilities to determine whether the decisions to disclose or not to disclose are ethically responsible.

Stakeholders

This section describes the stakeholders in the three cases under discussion. In the Stuxnet case, the stakeholders were Israel, Iran, the U.S., and the neighbouring countries. In the Osirak case, the stakeholders were Israel, Iraq, France, Italy, and the U.S.; while in the al-Kibar case, the stakeholders were Israel, Syria, and the neighbouring countries. **Table 4** depicts the stakeholders.

Stakeholders	Osirak	al-Kibar	Stuxnet
France	x		
Iran			x
Iraq	x		
Israel	x	x	x
Italy	x		
Syria		x	
U.S.	x		x
Neighbouring Countries		x	x

Table 4: The various stakeholders in the three cases

The Iranian government was and is continuing to pursue the development of nuclear weapons. The nuclear facility in Natanz shows their intent to enrich uranium; and, as long as the Iranians pursue this course of action, they will be perceived as an existential threat to Israel. In the al-Kibar-case, when al-Assad was working to repair relations with the West, Iran sent a government representative calling "Assad's plan 'unacceptable' and threatened that it would spell the end of the two countries' strategic alliance and a sharp decline in relations" (Follath & Stark 2009).

From Israel's perspective, countries "hostile to Israel and that call for its destruction must not be allowed to develop a nuclear military capability that could be used against Israel" (Yadlin 2018). This is also known as "the Begin Doctrine". Based on this doctrine, Israel conducted air strikes in Iraq and Syria. A different course of action was chosen against Iran: using cyber tools to affect the program.

Iraq was working to develop nuclear weapons in the 1970s (NTI 2019). It was cooperating with France and Italy to receive the necessary equipment for it.

Syria was developing a nuclear reactor covertly with the help of North Korea. The development of the reactor was done without the knowledge of Russia. After the nuclear program was destroyed, al-Assad was looking to repair relations with the West.

The U.S. condemned the attack on Osirak. Jeane J. Kirkpatrick, an American delegate to the United Nations, said that the US was '"shocked by the Israeli air strike, which exacerbated deeper antagonisms in the region'" (Nossiter 1981). Having said that, Kirkpatrick noted after a UN resolution that '"nothing in the resolution will affect my Government's commitment to Israel's security'" (Nossiter 1981). Now to the details of the "al-Kibar"-case. The U.S. was informed that al-Bashar was working with North Korea to develop a nuclear reactor in northern Syria. However, the CIA wanted to make sure that the threat was real; a harsh lesson was learned that Saddam Hussein had weapons of mass destruction (Follath & Stark 2009). In the Stuxnet case, the U.S. was unwilling to help Israel with a request on "bunker-busting bombs" (Sanger 2009). However, Israel had developed a cover program since 2008 to "penetrate Iran's nuclear supply chain abroad, along with new efforts, some of them experimental, to undermine electrical systems, computer systems and other networks on which Iran relies. It [was] aimed at delaying the day that Iran [could] produce the weapons-grade fuel and designs it need[ed] to produce a workable nuclear weapon" (Sanger 2009).

France condemned the attack. The French Foreign Ministry said the "main reactor, which uses highly enriched uranium fuel suitable for atomic weapons, was 'seriously damaged'" (Lewis 1981). Of the 150 French citizens working at the site, one technician was killed.

Italy "said that none of the 20 Italian technicians at the project had been injured" (Lewis 1981). They were working on the site to manage radioactive materials.

Neighbouring countries posed a threat to the operation. The designers of the air strike in Osirak noted this threat and plotted a dogleg course "to best avoid detection by Jordanian radar to the north and the Saudi E-3 Airborne Warning and Control System operating to the south" (Correll 2012, p. 61).

Technical and Professional Problems

This section depicts the technical/professional problems with the three cases. These problems are clustered in centrifuges, hardware, and software; fighter jet distance; and adversary technical threats.

Centrifuges, hardware, and software

The technical problem was to identify the type of centrifuge, the hardware, the software, and the related vulnerabilities to exploit. Then, a testbed had to be developed with the centrifuges, and related hardware, and software. In addition, the software that was going to exploit the vulnerabilities and degrade the uranium enrichment process had to be developed and tested.

Fighter jet distance

The al-Kibar case required the F-4 Phantoms and Skyhawks to fly over 1,000 miles. These fighter aircraft did not have the required capability. This technical problem was solved by acquiring the modern F16s. The professional problem to pilot the new F16s was solved by having the pilots train in the U.S.

Adversary technical threats

The technical threats consisted of radar systems, missile systems, and users. The threats stemming from radar and missile systems were managed as follows: avoid detection by plotting a "dogleg course" (Correll 2012, p. 61); using electronic warfare and "military computer hacking and electronic intelligence methods [and disabling] two radar systems" (Evans 2017). The users in the Stuxnet case were tricked by showing a picture of the sensors as if nothing were wrong with the centrifuges.

Ethical Issues

The ethical issues with these technologies are how they can be used by actors in each case to conduct air strikes, electronic warfare (EW), and offensive cyberspace operations, as well as to exploit zero-day vulnerabilities. These technologies are used by professionally trained pilots, sabotage units, EW-units, and cyberspace operators. Each capability has a controller, (for example, pilots, sabotage personnel, EW-personnel, and cyberspace operators, and the targets affected by the activities of pilots, sabotage personnel, EW-personnel, and cyberspace operators). The intentions of the pilots, sabotage personnel, EW-personnel, and cyberspace operators involve reasoning and establishing a purpose for the same, as well as a goal. The targets affected by the purpose of pilots, sabotage personnel, EW-personnel, and cyberspace operators involve the technical issues of bombs, ground attacks, and non-kinetic warfare (such as offensive cyberspace operations). The interaction between the technical use of pilots, sabotage units, EW, and cyberspace operations to destroy or to degrade nuclear programs and the technical effects of pilots, sabotage units, EW, and cyberspace operations as they impact nuclear programs and the people involved, are where ethical issues arise.

The Social Consequences of Vulnerabilities

Vulnerabilities and zero-days in information systems can have negative effects on society and on global relations. Information systems that can be exploited can be a part of an electric grid system, financial system, or transport system. Not disclosing identified vulnerabilities and zero-days in these systems pose the risk of adversaries' identifying these vulnerabilities and exploiting them for their own purposes. Affecting the electric grid system could lead to the residents of a city, or part of a city, being without electricity. Attacks on the financial system could make it difficult or impossible for companies to do business. Finally, attacking the transport system could interrupt just-in-time-deliveries to various stores, such as food stores.

Technical Conclusions

Air strikes and cyberspace operations were used to degrade and destroy nation states' nuclear programs posing an existential threat to Israel. The air strike in 1981 in Iraq received a great deal of international criticism and condemnation. Not only did the air strikes breach international law by flying into another country's territory, but they also led to the death of people and destruction of facilities. The 2007 air strike in Syria was hushed down by the Syrian government, which did not lead to any international criticism and condemnation.

The alleged U.S. and Israeli cyberspace operation never received the same media attention as the air strike operations. It also never received the attention of the users and decision makers in Iran working in their nuclear program. The ethical issue with neutralizing threats through air strikes and the likelihood of casualties show the level of risk that was involved. This can be compared to the ethical issue of not disclosing vulnerabilities and using exploits to degrade a nuclear program, which leads to a significantly lower level of risk, international condemnation, and casualties. Another major element here is the difficulty of attribution for attacks, such as the Stuxnet attack.

Policy-Related Ethical Analysis

The continual advances in technology result in society's greater dependency on this technology. At the same time, it is known that hardware and the software that operates that hardware contain vulnerabilities. It is, therefore, important to identify possible problems and attempt to anticipate ethical problems. Identifying possible problems and anticipating ethical problems can be used as the basis for policy development. Therefore, the vulnerability equity process was developed to assist decision makers determine whether to disclose or restrict the knowledge about zero-day vulnerabilities in hardware and software. Hardware and software consist of computer artefacts.

A group of scholars met to discuss "ethical guidance for the research and application of pervasive and autonomous information technology" (Miller 2011, p. 57). The resulting document was dubbed "Principles Governing Moral Responsibility for Computing Artifacts" (Miller 2011, p. 57) and consists of five rules about moral responsibility for computing artefacts. These rules will be used to develop policy recommendations. Moral responsibility is "that people are answerable for their behaviour when they produce or use computing artifacts" (Miller 2011, p. 57). The first rule states, "The people who design, develop, or deploy a computing artefact are morally responsible for that artefact, and for the foreseeable effects of that artefact. This responsibility is shared with other people who design, develop, deploy or knowingly use the artefact as part of a sociotechnical system" (p. 58). The application of this rule could be extended to those discovering zero-days.

Discovering zero-days involves actions performed on software and hardware to put the hardware and software in a state the developers never intended the software and hardware to be in. For example, the developer(s) of fingerd and sendmail in the 1988 Morris case never intended the software to receive more than 512 bytes of instructions. However, Morris took actions and found that by sending a 536-byte request he could execute code on the target system. The application of this rule for policy would be that those discovering zero-days should not disclose them in cases of existential threats where actions to manage the threats could lead to international condemnation.

The second rule of the five states,

> The shared responsibility of computing artefacts is not a zero-sum game. The responsibility of an individual is not reduced simply because more people become involved in designing, developing, deploying or using the artefact. Instead, a person's responsibility includes being answerable for the behaviours of the artefact and for the artefact's effects after deployment, to the degree to which these effects are reasonably foreseeable by that person. (Miller, 2011, p. 58)

This second rule could be applied to the discovery of zero-day vulnerabilities if it would state that those discovering zero-days are responsible for how that zero-day is exploited. Applying this to the al-Kibar and Osirak cases means that the pilots, sabotage units, and EW- operators are responsible for how they exploited their capabilities.

The fourth rule of the five rules is, “People who knowingly design, develop, deploy, or use a computing artefact can do so responsibly only when they make a reasonable effort to take into account the sociotechnical systems in which the artifact is embedded” (Miller 2011, p. 58). One application of this rule would be that those discovering a zero-day and who know it will be exploited to manage an existential threat should exploit it responsibly and should not disclose it.

Conclusions: Anticipatory Ethical Recommendations/Policy

Anticipatory ethics can be used to develop policy recommendations when it is used in conjunction with the conclusions of the preceding application of the five rules for computing artefacts. It is anticipated that existential threats from an adversary will always be met with a combination of kinetic and non-kinetic responses.

The response is the result of the adversary’s intent to gain capability once it is complete and, if used, poses an existential threat. The Stuxnet case shows how zero-day vulnerabilities in hardware and software were exploited to degrade the nuclear program of Iran. It did so ‘under the radar’, and it was not until the software came out of its target that it was revealed to the world. Even then, it never received any international condemnation because it was difficult to attribute it.

This would have never been possible if the exploited zero-day vulnerabilities had been disclosed. If these zero-day vulnerabilities had been disclosed, the first option, to use “bunker bombs” might have been applied. The results would have reflected the results of the Iraqi and Syrian cases, which included high levels of risk to pilots, sabotage units, EW-personnel, and personnel working at the nuclear program facilities, while also risking armed conflict with neighbouring countries.

The exploitation of zero-day vulnerabilities in the Stuxnet case removed all those risks. The lessons learned from the Stuxnet case give a good indication of what needs to be anticipated about future cases involving vulnerability disclosure.

Author’s Note

An earlier version of this paper was published in the *Proceedings of the 15th International Conference on Cyber Warfare and Security* 2020, eds. Brian K Payne & Hongyi Wu, Academic Conferences and Publishing International Limited, Reading, UK, pp. 254-61.

References

BBC 2006, ‘Factfile: How Osirak was bombed’, viewed 6 August 2020, <http://news.bbc.co.uk/2/hi/middle_east/5020778.stm>.

——2018, ‘Israel admits striking suspected Syrian nuclear reactor in 2007’, viewed 6 August 2020, <https://www.bbc.com/news/world-middle-east-43481803>.

Correll, JT 2012, *Air strike at Osirak*, *Air Force Magazine*, vol. 95, no. 4, pp. 58-62, viewed 6 August 2020, <https://www.airforcemag.com/PDF/MagazineArchive/Documents/2012/April%20 2012/0412osirak.pdf>

Evans, A 2017, ‘A lesson from the 1981 raid on Osirak’, viewed 6 August 2020, <https://www.wilsoncenter.org/blog-post/lesson-the-1981-raid-osirak>.

Follath, E & Stark, H 2009, ‘The story of “Operation Orchard”: How Israel destroyed Syria’s Al Kibar nuclear reactor’, *Der Spiegel*, viewed 6 August 2020, <https://www.spiegel.de/international/world/the-story-of-operation-orchard-how-israel-destroyed-syria-s-al-kibar-nuclear-reactor-a-658663.html>.

Furrow, D 2005, *Ethics: Key concepts in philosophy*, Bloomsbury Academic, London, UK.

Huskaj, G 2019, ‘The current state of research in offensive cyberspace operations’, *Proceedings of the 18th European Conference on Cyber Warfare and Security*, eds. T Cruz & P Simoes, Academic Conferences and Publishing International Ltd., pp. 660-7.

Langner, R 2013, ‘*To kill a centrifuge: A technical analysis of what Stuxnet’s creators tried to achieve*’, viewed 6 August 2020, <https://www.langner.com/wp-content/uploads/2017/03/to-kill-a-centrifuge.pdf>.

Lewis, P 1981, ‘France condemns attack and rejects Israeli account’, viewed 6 August 2020, <https://www.nytimes.com/1981/06/09/world/france-condemns-attack-and-rejects-israeli-account.html>.

Miller, KW 2011, ‘Moral responsibility for computing artifacts: “The rules”’, *IT Pro*, vol. 13, no. 3, pp. 57-9.

MITRE 2019, ‘CWE-20: Improper input validation’, viewed 6 August 2020, <http://cwe.mitre.org/data/definitions/20.html>.

Mortenson, MJ & Vidgen, R 2016, ‘A computational literature review of the technology acceptance model’, *International Journal of Information Management*, vol. 36, no. 6, pp., 1248-59, viewed 6 August 2020, <http://dx.doi.org/10.1016/j.ijinfomgt.2016.07.007>.

Murchu, LO 2010, ‘Stuxnet using three additional zero-day vulnerabilities’, viewed 2 March 2020, <https://www.symantec.com/connect/blogs/w32stuxnet-installation-details>.

Naraine, R 2010a, ‘Attack code published for unpatched Stuxnet vulnerability’, *zdnet.com*, viewed 6 August 2020, <https://www.zdnet.com/article/stuxnet-attackers-used-4-windows-zero-day-exploits/>.

——2010b, ‘Stuxnet attackers used 4 Windows zero-day exploits’, *zdnet.com*, viewed 6 August 2020, <https://www.zdnet.com/article/stuxnet-attackers-used-4-windows-zero-day-exploits/>.

Nossiter, BD 1981, ‘Israelis condemned by security council for attack on Iraq’, viewed 6 August 2020, <https://www.nytimes.com/1981/06/20/world/israelis-condemned-by-security-council-for-attack-on-iraq.html>.

NTI 2019, ‘Iraq’, viewed 6 August 2020, <https://www.nti.org/learn/countries/iraq/>.

——2020, 'Iran', viewed 6 August 2020, <https://www.nti.org/learn/countries/iran/nuclear/>.

NVD 2010a, 'CVE-2010-2568 detail', viewed 6 August 2020, <https://nvd.nist.gov/vuln/detail/CVE-2010-2568>.

——2010b, 'CVE-2010-2729 detail', viewed 6 August 2020, <https://nvd.nist.gov/vuln/detail/CVE-2010-2729>.

O'Leary, M 2019, *Cyber operations: Building, defending, and attacking modern computer networks*, 2nd ed, Apress, New York, NY, US.

Opall-Rome, B 2018, 'Declassified: How an Israeli operation derailed Syria's nuclear weapons drive', viewed 6 August 2020, <https://www.defensenews.com/global/mideast-africa/2018/03/20/just-declassified-how-an-israeli-operation-derailed-syrias-nuclear-weapons-drive/>.

Pfleeger, CP & Pfleeger, SL 2012, *Analyzing computer security*, 1st edn., Prentice Hall, New York, NY, US.

Sanger, DE 2009, 'U.S. rejected aid for Israeli raid on Iranian nuclear site', viewed 6 August 2020, <https://www.nytimes.com/2009/01/11/washington/11iran.html>.

Shipler, DK 1981, 'Israeli jets destroy Iraqi atomic reactor; Attack condemned by U.S. and Arab nations', viewed 6 August 2020, <https://www.nytimes.com/1981/06/09/world/israeli-jets-destroy-iraqi-atomic-reactor-attack-condemned-us-arab-nations.html>.

TOI STAFF 2019, '38 years later, pilots recall how Iran inadvertently enabled Osiraq reactor raid', viewed 6 August 2020, <https://www.timesofisrael.com/38-years-later-pilots-recall-how-iran-inadvertently-enabled-osiraq-reactor-raid/>.

WebDEViL 2010, 'Microsoft Windows - Task scheduler privilege escalation', viewed 6 August 2020, <https://www.exploit-db.com/exploits/15589>.

White House 2017, *Vulnerabilities Equities Policy and Process for the United States Government*, viewed 6 August 2020, <https://www.whitehouse.gov/sites/whitehouse.gov/files/images/External%20%20Unclassified%20VEP%20Charter%20FINAL.PDF>.

Xmco 2011, '*ACTUSÉCU 27*', viewed 6 August 2020, <https://www.xmco.fr/actu-secu/XMCO-ActuSecu-27-STUXNET_EN.pdf>.

Yadlin, A 2018, 'The Begin Doctrine: The lessons of Osirak and Dear ez-Zor', viewed 6 August 2020, <https://www.inss.org.il/publication/the-begin-doctrine-the-lessons-of-osirak-and-deir-ez-zor/>.

Fitting the Artificial Intelligence Approach to Problems in DoD

JS Hurley

College of Information and Cyberspace
National Defense University
Washington, D.C., United States

Email: john.hurley@ndu.edu

Abstract: *Emerging and disruptive technologies, due to advances made over the last couple of decades, have become the centrepiece of Department of Defense (DoD) concerns about national security. The technologies are unique because they can both benefit and hinder the DoD from its mission. Artificial Intelligence (AI) is revered in DoD circles as one of the most important of these technologies because of its potential as an absolute 'gamechanger' in cybersecurity operations. In this study, the focus is on the DoD fitting the Artificial Intelligence approach to its problems in a time of limited and diminishing resources.*

Keywords: *Artificial Intelligence, National Cyber Strategy, Cyber Conflict, DoD, National Security, Emerging and Disruptive Technologies*

Introduction

The potential transformative impact of AI across numerous DoD functions is implied in the current U.S. national strategy for Artificial Intelligence (The White House 2019). DoD, in its role as the 'steward of prosperity and security for the American public' must balance its efforts of unpredictable, committed, well-financed, and, in some cases, sophisticated groups of adversaries against the ethical, societal, and technical concerns about AI held by many segments of society. Of particular concern is the recognition by other actors, including major adversaries such as China and Russia, of the potential of AI to advance their global agendas. Such potential outcomes have created a new 'arms' race between a wide range of actors who recognize that AI can be a potential gamechanger in cyber operations. Significant concern also exists on the level of investments by chief competitors in AI and the fallout if the U.S. is not at least comparable in its investments with its chief competitors (Department of Defense 2018).

Some of the perceived potential benefits of AI to DoD include implementing predictive maintenance and supply, improving decision making and situational awareness, increasing the safety of operating equipment, and streamlining business process and prototypes. In addition, AI is expected to introduce new ways of working, including allowing components of the DoD to pass tedious physical and cognitive tasks on to machines. Some other highly anticipated capabilities of AI for DoD include extending adaptable problem solving and increasing the speed of delivery and rate of experimentation (The National Security Commission on Artificial Intelligence (NSCAI) 2020).

The potential benefits of AI, however, must be judged alongside a number of potential hurdles. For example, AI requires a highly skilled and specialized workforce because of the demands of an extensive technical background. In addition, AI requires a significant upgrade to the current DoD environment, which lacks sufficient and appropriate infrastructure. The infrastructure is critical if the DoD hopes to successfully deploy scalable neural network algorithms and to manage certain resources, such as the density of data handled, high performance computing networks, and required storage computing capacity. There is also a great deal of discontent among segments of society with respect to how ethical, social, and technical issues of AI are being handled by the federal government in general and by the DoD in particular. There is a major mistrust regarding how AI might be used in regard to decision making, especially when human lives and livelihoods may be at risk (Polonski 2018). Though not discussed very openly, there is also significant concern for the ways executives might use AI within the workplace to replace humans with machines. There is also a major lack of confidence and assurance in the predictability of AI in certain problems to be solved (Office of Science and Technology Policy 2019). For example, problems in which well-formed objectives are not determined *a priori* pose a challenge for Artificial Intelligence. In real-world cases, this is actually more of the norm than the exception, and determining these objectives presents a formidable obstacle (Lipton 2020). The aforementioned AI benefits and hurdles are a small sample of a larger pool of factors that can directly impact a number of different functions and operations in the DoD in which Artificial Intelligence is believed to be applicable.

There unfortunately continues to remain a lack of a consensus on the definition of cyberwarfare around the globe. In addition, there are increasing challenges posed by a wide variety of actors (such as individuals, groups, and nation states) below the level of armed conflict that have put a lot of pressure on the DoD in terms of what qualifies as cyberwarfare. Over the last few years, increasing threats and attacks to the peace and prosperity of the United States have been revealed in conflicts facilitated in the cyber domain. The White House has developed a national cyber strategy that notes its top priorities and pillars relevant to national security. Critical to the success of these endeavours is ensuring that cyberspace is secured. Unfortunately, advancements in computing and information technologies have led to an increase in the sophistication and frequency of malicious cyber activities. To address many of the aforementioned concerns, a National Cyber Strategy was developed from *Executive Order 13800, Strengthening the Cybersecurity of Federal Networks and Critical Infrastructure*. The National Cyber Strategy has four fundamental pillars:

- To protect American people, the homeland, and the American way of life;
- To expand American influence abroad to extend the key tenets of an open, interoperable, reliable, and secure Internet;
- To preserve peace and security by strengthening the ability of the United States—in concert with allies and partners—to deter and, if necessary, punish those who use cyber tools for malicious purposes; and
- To promote American prosperity by nurturing a secure, thriving digital economy and fostering strong domestic innovation (The White House 2018).

However, given the state of economic upheaval due to numerous factors such as the current global coronavirus pandemic, competing interests within the federal government, continuing surveillance, and attempts at disruption by adversaries, it is in the interest of the DoD to prioritize its use of AI to ensure that it gets the best Return On Investment (ROI). The U.S. economy is already being adversely affected by the COVID-19 pandemic; for example, both domestic and global supply chains are disrupted; businesses are shuttered; and interstate commerce is restricted. As should be expected, the DoD is not immune to the economic fallout associated with competing interests, as well as by the coronavirus pandemic. Hence, it is prudent for the DoD to be fiscally astute, cognizant of the impending financial challenges, and able to find compelling efficiencies (Egel *et al.* 2020).

Given the benefits and hurdles that AI can present, this study will focus on how the DoD can fit the AI approach to problems in lieu of budget uncertainties. Though a previous effort focused on the use of AI in cyberwarfare, this study will not focus on cyberwarfare because there still remains a lack of consensus in its definition (Hurley & Potter 2020). Also, the scope of this study is limited to cyber events below the level of armed conflict—deemed in this study to be ‘cyber conflicts’. In the first section of the paper, the focus will be on some of the benefits of AI that the DoD sees for itself and the nation. Next, focus shifts to some of the challenges that the DoD faces in its implementation of Artificial Intelligence. The next section discusses the results of a prior study on DoD decision making and how to prioritize the DoD cyber priorities and best align them with specific AI attributes. It is the expectation that such results will enable the DoD to optimize its use of AI in its functional requirements.

The Benefits that DoD Sees in AI to Itself and to the Nation

The market analysis firm, IDC, predicts that the estimated sum of the world’s collective data, for example, 33 ZBs will grow to almost 175 ZBs by 2025 (Reinsel, Gantz & Rydning 2018). One of the most important goals of the DoD is to be more efficient in its delivery of strategy, operations, and tactics. However, impeding this goal are two very important factors. One is that the DoD is an organization mired in a traditional culture that has been data averse. Additionally, the DoD is well known as a siloed data environment, in which, at times, mission effectiveness is limited, critical decision points are left inadequately addressed, and cooperation is inhibited (Lyle 2019). The DoD has failed to fully appreciate how data can better enable decisions and, ultimately, outcomes. However, now the Pentagon is being forced to change its view of data because an enormous amount of structured and unstructured data that the DoD requires are generated by a variety of different sources including automated cybersecurity systems, drones, terrorist databases, the Internet of Things (IoTs), and other sources (Costlow 2014). Additionally, military and intelligence applications are increasingly relying on huge data pipelines to drive mission-critical decision making and intelligence gathering. The speed at which the warfighter is able to analyse, collect, process, and understand data directly impacts mission success and warfighter survivability (Whaley 2019). Hence, the DoD must take a more data-based approach to decision making in which it quickly analyses and utilizes big data so that it can address requirements in a desired timeframe.

Artificial Intelligence has been proposed as a major emerging and disruptive technology for enhancing the DoD’s ability to meet a number of its objectives, especially its need to simplify data workflows and to improve the accuracy and speed at which repetitive tasks get completed. Recent successes and advancements of AI have prompted a surge in the push for increased automation in DoD functions. However, there remains quite some scepticism due to the unpredictability of AI that still requires some level of human intervention, when high-level reasoning and judgment are required. It is believed that AI could create critical operational and strategic advantages if it can eventually be scaled across the DoD enterprise. The scaling of AI has been inhibited largely by organizations that are only either piloting AI or using it in a single business process, thus gaining only incremental benefits (Fountaine, McCarthy & Saleh 2019). Organizational and cultural changes that are required are not receiving the attention, resources, and time required to bring AI to a level of scale capable of delivering meaningful value to the DoD. For AI to reach the level of scale needed, pilots across an organization must be produced in a fast, consistent, and repeatable manner. In addition, the pilots must enjoy widespread end-user adoption. Ultimately, the DoD must establish key AI standards and building blocks; attract and develop AI talent; introduce new operational models; and identify and implement new organizational approaches. If these things can happen, then the DoD likely could take advantage of AI systematically across its entire enterprise. The DoD can learn a valuable lesson from industry, which has been much more successful in scaling AI up to a level at which it has gained meaningful value. Companies that have been successful in scaling AI up to a meaningful level seem to have three common processes in place:

1. The ability to transition to more of an interdisciplinary collaborative environment from the traditional siloed work environment that the DoD is well known to facilitate. In a collaborative environment, there is the benefit of different approaches and perspectives that can expand the viewpoints to ensure that initiatives address broad organizational priorities;
2. The option to switch to a more evidence-based decision-making approach, rather than the traditional autocratic, experience-based, leader-driven decision-making approach that has dominated the DoD throughout its history. In this case, the DoD would merge the best available evidence with critical thinking and varied (sometimes diverging) perspectives to arrive at an optimum decision; and finally,
3. The ability to transition to an adaptable, agile, flexible mindset instead of the historical rigid and risk-averse way of thinking that has typically dominated DoD decision making. In this new way of thinking, decisions and results can be obtained in weeks rather than in months (Fountaine, McCarthy & Saleh 2019).

Artificial Intelligence is proposed to enable a stronger defense of the nation’s critical infrastructure and to protect the security and safety of U.S. citizens. In particular, AI is projected to enhance the DoD’s ability to identify, predict, and respond to physical and cyber threats from various sources; to discourage attempts to disrupt U.S. critical infrastructure; and to strengthen the defense of the homeland from attacks. The U.S. must adopt AI to maintain strategic positions which enable it to prevail on future battlefields and use and develop AI technologies in ways that advance peace, security, and stability in the future. As it relates to cybersecurity, hackers need only to modify small segments of their code to circumvent many of the existing defenses, because conventional cybersecurity tools look only for historical matches to known malicious code. AI-enabled tools, on the other hand, can present a more dynamic and comprehensive barrier to attack because they can be trained to detect anomalies in broader patterns of network activity (Theohary 2018; Sayler & Harris 2020). In a Defense Advanced Research Project Agency (DARPA) 2016 Cyber Grand Challenge, the potential power of AI-enabled cyber tools was demonstrated. Within the challenge, two distinct advantages were shown: first, in the potential ability of a singular algorithm to play offense and defense simultaneously and, second, in the potential speed at which AI-enabled cyber tools can produce results—both with potential major upsides for future cyber operations.

Although the DoD is considering a number of diverse AI applications across its different levels, it is currently leaving Research and Development (R&D) efforts to the discretion of the Defense Advanced Research Project Agency (DARPA), the Intelligence Advanced Research Project Agency (IARPA), and the R&D organizations within the individual services. Additionally, the individual services and the DoD components are required to coordinate with the Joint Artificial Intelligence Center (JAIC) on any AI initiatives that cost over $15 M annually (Shanahan 2018). The JAIC has also been tasked with projects that leverage AI to address pressing operational challenges in the National Mission Initiatives (NMI). It is anticipated that AI can be incorporated into operations and decision making to generate military advantage, as well as to reduce the risk to fielded forces. As a major aid to service members, AI is expected to improve readiness, reduce operational costs, and help better maintain key equipment. It is projected that AI may also enhance the ability to implement the Law of War, which is a component of international law that regulates the conduct of warring parties and the conditions for war (Preston 2015). By enhancing mission precision and improving the accuracy of military assessments, AI can reduce the risk of casualties to military personnel and civilians, as well as prevent other collateral damage. In particular, AI is projected to equip, organize, and train forces to prevail in a world in which military systems empowered by AI are prominent in all domains (Morgan *et al.* 2018). The DoD is developing AI applications for a range of military functions. For example, AI research is underway in the fields of analysis, command and control, cyber operations, Information Operations (IO) intelligence collection; logistics; as well as a variety of autonomous and semiautonomous vehicles. AI has already been incorporated into military operations in Iraq and Syria (Congressional Research Service 2020).

There are expectations that AI will be particularly useful in intelligence, due to the large data sets available for analysis. As noted earlier, the JAIC will play a critical role in the coordination of DoD efforts to develop, mature, and transition Artificial Intelligence technologies into operational use. In particular, the JAIC is spearheading Project Maven to incorporate computer vision and machine-learning algorithms into intelligence collection cells that would comb through footage from uninhabited aerial vehicles and then automatically identify hostile activity for targeting. Currently, Project Maven is using AI algorithms to identify insurgent targets in Iraq and Syria. In this capacity, AI is intended to automate the work of human analysts who currently spend hours sifting through drone footage for actionable information, which may potentially free analysts to make more efficient and timely decisions based on the data (The Congressional Research Service (CRS) 2020). The aforementioned projects are just a sample of a number AI projects underway that support U.S. citizens and service members around the globe. However, as there are benefits of Artificial Intelligence that the DoD can utilize, there are also potential challenges of AI and its implementation that must also be considered, as shown below.

Challenges of Artificial Intelligence that DoD Must Address

It is no secret that the DoD has been trying to achieve a sensitive, yet important balance between the use of Artificial Intelligence to meet its mission against concerns about how AI will be used efficiently and ethically. This has been visibly apparent in conflicts exhibited in the pushback from the private sector and the general public (Wirtz, Weyerer & Geyer 2019). Mainstream corporations like Google have seen major resistance within their rank and file on the use of emerging and disruptive technologies, like Artificial Intelligence in weapon systems and for offensive purposes (Masunaga 2018). This was so much of an issue that Google declined to renew the Pentagon's

Artificial Intelligence drone deal after receiving major backlash from some of its employees and customers (Bergen 2018). There are, as well, additional concerns that AI may be unpredictable or vulnerable to unique forms of manipulation. In addition, there exist a number of other challenges related to culture, personnel, processes, and technology that must be addressed if the adoption of AI by the military is to be comprehensive.

Most of the development of AI is taking place in the commercial sector and developed primarily for civilian requirements. Unfortunately, the civilian requirements create unique challenges for AI integration into the military environments, which tend to be more complex, unique, and designed specifically for combat and adversarial conditions. For example, commercial semi-autonomous vehicles have largely been developed in and for data-rich environments with reliable GPS resources, comprehensive terrain mapping, and up-to-date information on traffic and weather conditions obtained from other networked vehicles (Cainis 2017). On the other hand, such a vehicle developed for the military would need to be able to operate in locations which demand robust GPS devices that can minimize potential jamming by an adversary. In addition, autonomous or semi-autonomous military ground vehicles would likely have to negotiate through dangerous and rough terrain (CRS 2020).

Additionally, there is also a lot of tension between the DoD and companies on recruiting and retaining personnel with expertise in AI. The simple reality is that the federal government has not been able to compete with industry on the benefits and salaries offered; and, as a result, the federal government has had a hard time attracting and retaining skilled personnel with strong AI backgrounds (National Security Commission on Artificial Intelligence (NSCAI) 2019). There is a keen awareness that salaries and research funding in the federal government significantly lag behind that of commercial companies. Two efforts, in particular, are noted to possibly address some of the imbalance. In 2015, the Defense Digital Service was designed to recruit members from the commercial sector for one- to two-year assignments. Similarly, there is an effort known as the ‘AI Training Corps’ in which pay for advanced technical education is provided in exchange for two days a month of training with government systems and two weeks a year for major exercises.

There has also been a lack of comfortability with AI, both outside and inside of the DoD, regarding how AI technology is incorporated into the DoD’s processes and approaches. Current DoD processes, including those related to standards of acquisitions, data rights, Intellectual Property (IP), performance, and safety present additional challenges to the transition of AI solutions from civilian to military environments. Military and civilian standards of safety and performance are often not well aligned to support an easy transfer between them. An acceptable civilian AI failure rate may be well outside of acceptable tolerances in a combat environment. In addition, the DoD has been immersed in a constant battle with the commercial sector with regard to the Intellectual Property (IP) of the companies and how the DoD will handle IP. Companies have been resistant to revealing certain information to the DoD because of concerns over what and how information may be shared by the DoD, possibly compromising some competitive advantage. As a matter of fact, in 2017, a Government Accountability Office (GAO) report noted heightened concern by industry due to increased pressure by the DoD for unlimited technical data and software rights or government purpose rights rather than limited or restricted rights (Government Accountability Office 2017). In the next section, the results of a previous study that explored cyber problems at the DoD in an AI-based framework are discussed and incorporated into this present effort.

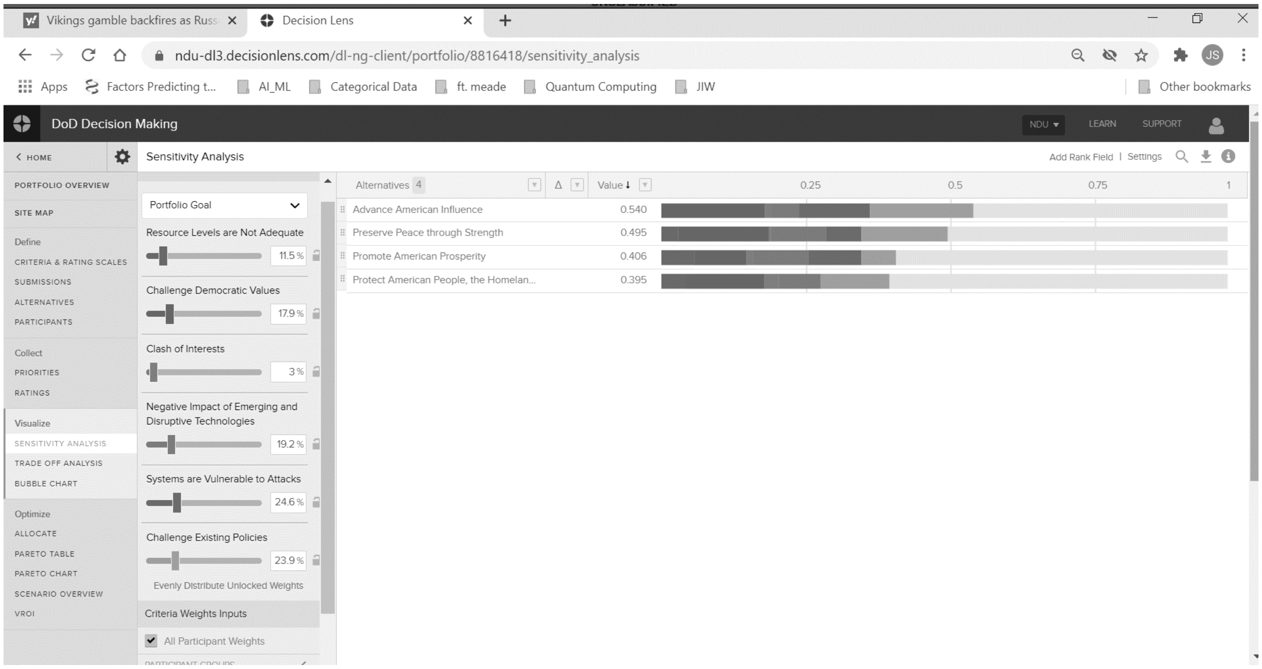
Methods

For this study, the Analytic Hierarchy Process (AHP) is used to assess the impact of AI in the DoD cyber problem space. The AHP captures strategic goals as a set of weighted criteria. It is a method used to select and to prioritize projects (Saaty 2012) through a pairwise comparison (a comparison of elements in pairs) against given criterion. The decision-making platform used in this study is based upon the AHP and incorporates an impact rating scale with the following weights: 1 (strong positive impact); 0.75 (moderate positive impact); 0.5 (minimum positive impact); 0.25 (neutral or no impact); and 0 (negative impact). It is acknowledged that a different value for a negative impact could be employed. However, for this effort, the impact rating scale set by the platform, as noted above, is used.

Discussion and Results

Figure 1, below, represents some of the factors that the DoD should consider in its decision making to address cyberwarfare requirements (Hurley & Potter 2020). The illustration (**Figure 1**) identifies the vulnerability of information assets (for example, systems, information, networks, and data) to attacks as the top priority of the DoD in the context of cyberwarfare. The inherent synergy between cyberwarfare and cyber conflicts is a natural incentive to use either term. However, the author reiterates that the focus of this study will be on cyber conflicts instead of cyberwarfare because of the lack of a consensus on the definition of cyberwarfare and the focus on events below the level of armed conflict. The challenges of existing policies often locked into cultural inflexibility and tradition make it very difficult for organizations like the DoD to change how decisions are made at the senior levels. One of the major misconceptions that such organizations have is that they get too caught up in the technology. As a result, they fail to realize that a purely IT—or technology—centred approach will be unsuccessful. The failure is largely due to the fact that such approaches do little if anything to truly address the culture and tradition challenges linked more to people’s intricate beliefs and behaviours and information management than with information technology. Organizations must develop a comprehensive and balanced information-oriented framework, which recognizes and implements synergies across three information capabilities, including: peoples’ values and behaviours pertaining to the use of information; information technology practices; and information management practices (Marchand, Kettinger & Rollins 2000). What is at odds here is that when huge organizations like the DoD focus so much on the technology, they often fail to address the value of information to people and their subsequent behaviours and motivations, which are necessary if a true transformation is to be realized.

Figure 2, below, represents some of the general ways that AI could be used by the DoD in lieu of the four pillars of the national cyber strategy. It appears that problem areas that focus on securing networks, systems, information and data (97.0%); securing critical infrastructure (94.3%); and building technical foundation of trust, security, and reliability (93.4%) represent the functions the DoD should most likely concentrate on in its use of AI. It is also important to acknowledge the closeness of about 1% between the weighting of securing the critical infrastructure and building technical foundation of trust, security, and reliability because additional or different voting could sway the weighting either way. Some flexibility is needed in this case in terms of which truly has the highest priority, given that they are so close (within 1% of each other). However, clearly it appears that each of the above top three DoD AI-focused areas share the common thread of security. **Table 1**, also below, represents a breakdown of the AI attributes in terms of benefits and hurdles which help to better realize those features of AI that are better suited for the most heavily weighted DoD cyber problem areas, listed earlier.



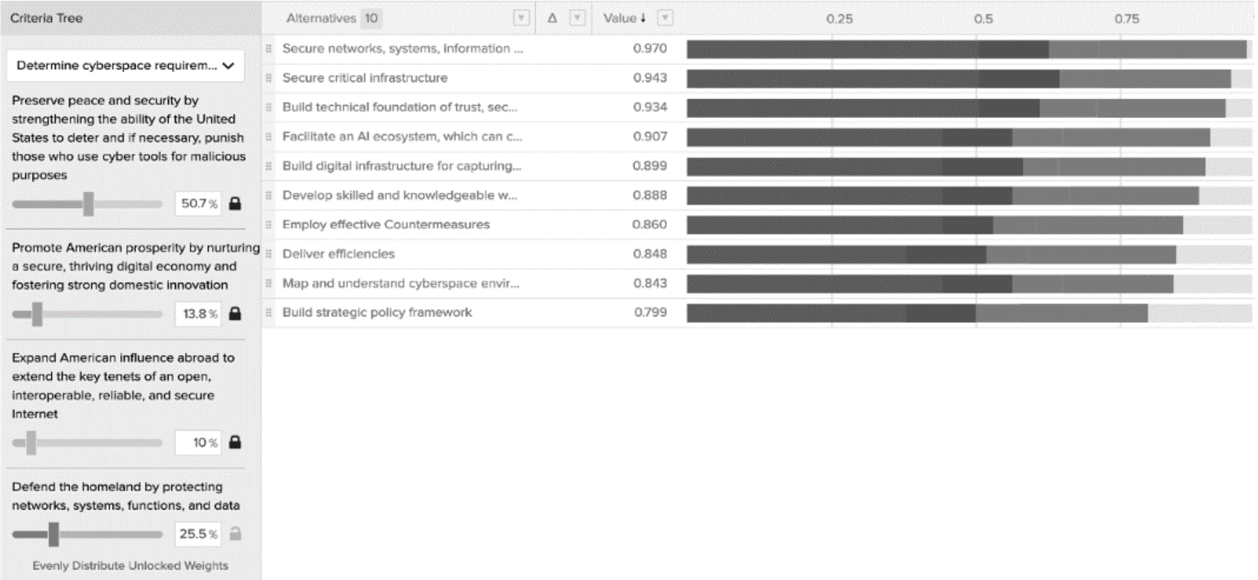
Portfolio Goal

Resource Levels are Not Adequate	11.5 %
Challenge Democratic Values	17.9 %
Clash of Interests	3 %
Negative Impact of Emerging and Disruptive Technologies	19.2 %
Systems are Vulnerable to Attacks	24.6 %
Challenge Existing Policies	23.9 %

☑ All Participant Weights

Alternatives 4	Value
Advance American Influence	0.540
Preserve Peace through Strength	0.495
Promote American Prosperity	0.406
Protect American People, the Homelan...	0.395

Figure 1: Important factors that the DoD needs to consider in AI addressing cyberwarfare



Determine cyberspace requirem...

Preserve peace and security by strengthening the ability of the United States to deter and if necessary, punish those who use cyber tools for malicious purposes	50.7 %
Promote American prosperity by nurturing a secure, thriving digital economy and fostering strong domestic innovation	13.8 %
Expand American influence abroad to extend the key tenets of an open, interoperable, reliable, and secure Internet	10 %
Defend the homeland by protecting networks, systems, functions, and data	25.5 %

Alternatives 10	Value
Secure networks, systems, information ...	0.970
Secure critical infrastructure	0.943
Build technical foundation of trust, sec...	0.934
Facilitate an AI ecosystem, which can c...	0.907
Build digital infrastructure for capturing...	0.899
Develop skilled and knowledgeable w...	0.888
Employ effective Countermeasures	0.860
Deliver efficiencies	0.848
Map and understand cyberspace envir...	0.843
Build strategic policy framework	0.799

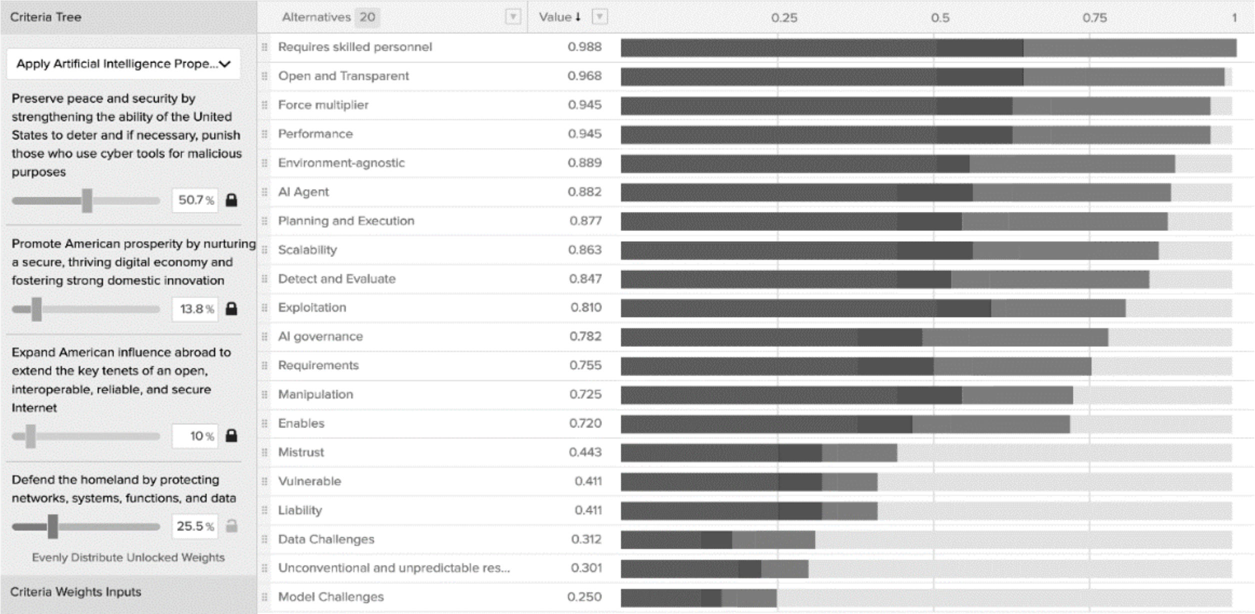
Figure 2: AI desired uses in DoD cyber conflicts

It is important to reinforce the need to maintain a consistent baseline in the weighting across the DoD cyber problems under consideration. The four pillars, prioritized in terms of weighting "preserve peace and security by strengthening the ability of the United States—in concert with allies and partners—to deter and, if necessary, punish those who use cyber tools" (50.7%); "expand American influence abroad to extend the key tenets of an open, interoperable, reliable, and secure Internet" (10%); "promote American prosperity by nurturing a secure, thriving digital economy and fostering strong domestic innovation" (13.8%); "defend the homeland by protecting the information assets" (25.5%) were consistently set (normalized) across each of the profiles as represented in **Figures 2-4**.

It is important to acknowledge that in the data platform used in this study, the four pillars corresponded to the criteria whereas the specific areas of the DoD cyber focus and AI attributes were respectively shown as alternatives or options. Although this information is useful and provides some insight that is helpful, to actually fit the AI approaches to a DoD problem area, a more granular view of the AI attributes is needed. **Figure 3** represents the AI attributes against the four pillars with a ranking and prioritization of the AI attributes. It is noted that the requirement for skilled personnel is, as might be expected, near the top of the DoD's requirement listing.

Benefits	**Hurdles**
Secure networks, systems, information, and data	Requires skilled personnel
Facilitate AI ecosystem	Open and transparent
Secure critical infrastructure	Build technical foundation
Deliver efficiencies	Scalability
Build digital infrastructure	Build strategic policy framework
Act as a Force Multiplier	AI governance
Perform	Clear requirements
Map and understand cyber environment	Manipulation
Support planning and execution	Exploitation
AI Agent perceives and acts upon environment	Mistrust
Employ effective countermeasures	Vulnerability
Detect and evaluate	Liability
Environment-agnostic	Data challenges
Enables	Unconventional and unpredictable results
Develop skilled and knowledgeable workforce	Model challenges

Table 1: AI benefits and hurdles


Criteria Tree
Apply Artificial Intelligence Prope...
Evenly Distribute Unlocked Weights
Criteria Weights Inputs
Alternatives 20
Value
0.25
0.5
0.75
1

Preserve peace and security by strengthening the ability of the United States to deter and if necessary, punish those who use cyber tools for malicious purposes: 50.7 %

Promote American prosperity by nurturing a secure, thriving digital economy and fostering strong domestic innovation: 13.8 %

Expand American influence abroad to extend the key tenets of an open, interoperable, reliable, and secure Internet: 10 %

Defend the homeland by protecting networks, systems, functions, and data: 25.5 %

Alternatives	Value
Requires skilled personnel	0.988
Open and Transparent	0.968
Force multiplier	0.945
Performance	0.945
Environment-agnostic	0.889
AI Agent	0.882
Planning and Execution	0.877
Scalability	0.863
Detect and Evaluate	0.847
Exploitation	0.810
AI governance	0.782
Requirements	0.755
Manipulation	0.725
Enables	0.720
Mistrust	0.443
Vulnerable	0.411
Liability	0.411
Data Challenges	0.312
Unconventional and unpredictable res...	0.301
Model Challenges	0.250

Figure 3: AI benefits and hurdles

The AI benefits and hurdles and requirements that the DoD is considering are represented in **Table 2** and **Table 3**, below. In addition, in **Table 2**, it is also noted that the status of the AI attributes are also listed and assigned the label of either readiness or unreadiness. This assignment takes into consideration the earlier premise that not all AI attributes are mature enough for the DoD to use.

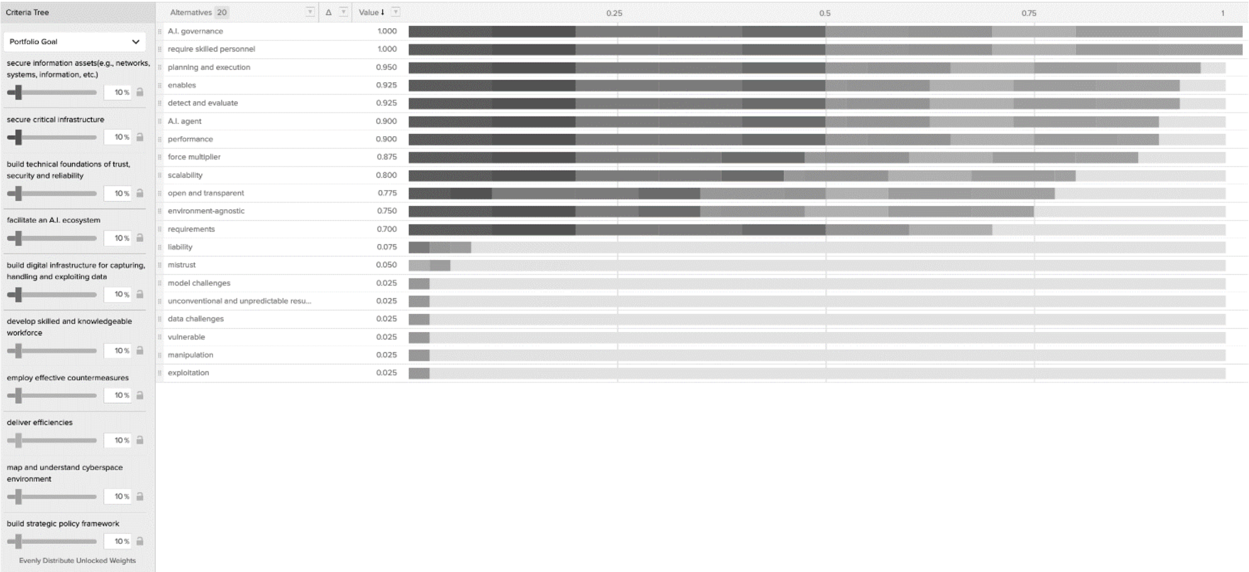
	AI Benefits and Hurdles and Status on Readiness	Weighting (%)	Status
1	sufficient specialized skilled personnel	98.8	unreadiness
2	open and transparent	96.8	unreadiness
3	act as a force multiplier	94.5	readiness
4	perform as intended	94.5	unreadiness
5	environment-agnostic	88.9	readiness
6	AI agent perceives and acts upon environment	88.2	readiness
7	support planning and execution	87.7	readiness
8	scalability	86.3	unreadiness
9	detect and evaluate	84.7	readiness
10	cannot be exploited	81	unreadiness
11	satisfactory AI governance	78.2	unreadiness
12	clear requirements	75.5	unreadiness
13	not be manipulated	72.5	unreadiness
14	enabler	72	readiness
15	trusted technology	44.3	unreadiness
16	lack vulnerability	41.1	unreadiness
17	reliability	41.1	unreadiness
18	data risks	31.2	readiness
19	generate unconventional and unpredictable results	30.1	readiness
20	model transparency	25	unreadiness

Table 2: AI benefits and hurdles and status on readiness

	DoD Requirements for AI	Weighting (%)
1	Secure information assets (networks, systems, information, for example)	97
2	Secure critical infrastructure	94.3
3	Build technical foundation of trust, security, and reliability	93.4
4	Facilitate an AI ecosystem	90.7
5	Build digital infrastructure	89.9
6	Develop skilled and knowledgeable workforce	88.8
7	Employ efficient countermeasures	86
8	Deliver efficiencies	84.8
9	Map and understand cyberspace environment	84.3
10	Build strategic policy forward	79.9

Table 3: Weighted DoD requirements for AI

Now, attention is focused specifically on the cyber conflicts that the DoD needs to address through the lens of AI benefits and hurdles. In looking to fit an approach to a problem, it is crucial to first define and understand the problem. Next, it is then important to identify and parse the approach in terms of its benefits and limitations and then relate them back to the deliverables of the problem. Additionally, it is essential to ensure that weightings are incorporated equally as measures to ensure that there is a common baseline that supports an 'apples-to-apples comparison'. **Figure 4**, below, represents the weighted and prioritized attributes of AI, as noted earlier.



secure information assets(e.g., networks, systems, information, etc.) 10 %
secure critical infrastructure 10 %
build technical foundations of trust, security and reliability 10 %
facilitate an A.I. ecosystem 10 %
build digital infrastructure for capturing, handling and exploiting data 10 %
develop skilled and knowledgeable workforce 10 %
employ effective countermeasures 10 %
deliver efficiencies 10 %
map and understand cyberspace environment 10 %
build strategic policy framework 10 %

Alternatives 20	Value
A.I. governance	1.000
require skilled personnel	1.000
planning and execution	0.950
enables	0.925
detect and evaluate	0.925
A.I. agent	0.900
performance	0.900
force multiplier	0.875
scalability	0.800
open and transparent	0.775
environment-agnostic	0.750
requirements	0.700
liability	0.075
mistrust	0.050
model challenges	0.025
unconventional and unpredictable resu...	0.025
data challenges	0.025
vulnerable	0.025
manipulation	0.025
exploitation	0.025

Figure 4: AI attributes (benefits and hurdles) and requirements

The lack of skilled and knowledgeable technical talent again is crucial to the DoD achieving any level of success. Whether that talent is home-grown or external to the DoD, knowledgeable and skilled workers are essential. So it comes as little surprise that the AI attributes in **Figure 4**, representing the requirement for knowledgeable and skilled personnel, is high on the list. However, it was a bit surprising to see AI governance on par with the requirements for skilled personnel. AI governance, although difficult to define, encompasses the managing of people and processes to get optimum results, as well as the monitoring and evaluating algorithms for bias, effectiveness, risk, and Return On Investment (ROI) (Taulli 2020).

Conclusion

In the federal fiscal year 2019, the defense spending budget totaled $676 B. However, the potential budget losses to the DoD due to COVID-19 could be comparable to a second sequestration. Hence, the DoD must be prudent in terms of how it invests anticipated reduced funding to meet its mission priorities, amongst them addressing “cyber conflicts” (Egel *et al.* 2020). In this study, the results showed that the top AI attributes that the DoD should focus on in its efforts to address cyber conflicts are AI governance and having skilled and knowledgeable workers. It was also noted that it is very important that huge global organizations like the DoD that seek to incorporate emerging and disruptive technologies into their solution portfolio, embrace a strategy that incorporates a balanced information-oriented framework. One of the biggest problems strategically is that the DoD does not give enough attention to people’s perspectives (internally and externally to the DoD) and information management as it seeks to deploy technology-centred solutions. This misstep is considered critical because it has been noted often that the DoD possesses an esteemed, yet exclusive, culture and tradition that drive much of its approaches to problem solving. As a result, it is critical that the DoD strategy reflect the view of people, information, and IT. Otherwise, organizations like the DoD that are mired in tradition and culture rarely move and improve in their decision making as much as they can or need to.

In this study, to get to the desired results, that is fit the AI approach to a DoD problem, it was necessary to first identify the problem and then identify and parse the AI approach in terms of its benefits and limitations. Next it was necessary to establish a common baseline in terms of weighting in the general problem areas (four pillars) against the AI attributes. In this effort, the general problem areas were defined in the decision platform as criteria. The four pillars were then considered in terms of more specific problems areas. Similarly, the AI attributes were also analysed into benefits and hurdles and defined within the platform as alternatives. Finally, the specific AI attributes were considered against the DoD-specific problem areas to obtain the top AI approaches relevant to the cyber conflict problem space.

Acknowledgements

Special thanks are extended to C Gebele, K Religioso, and S Warner at Decision Lens for their assistance and support with the decision-making platform used in this study.

Disclaimer

The opinions, conclusions, and recommendations expressed or implied are the author’s and do not necessarily reflect the views of the Department of Defense, any other agency of the U.S. Federal Government, or any other organization.

References

Bergen, M 2018, ‘Google won’t renew Pentagon Artificial Intelligence drone deal after staff backlash', *Los Angeles Times*, 1 June, viewed 30 October 2020, <https://www.latimes.com/business/la-fi-google-military-drone-20180601-story.html>.

Cainis, B 2017, 'Autonomous vehicles: Emerging policy issues', *CRS in Focus*, Washington, DC, US.

The Congressional Research Service (CRS) 2020, *Artificial Intelligence and National Security*, 26 August, viewed on 30 October 2020, <https://crsreports.congress.gov/product/pdf/R/R45178he>.

Costlow, T 2014, How Big Data is paying off for DOD, Defense Systems, Washington, DC, US.

Department of Defense 2018, 'Artificial Intelligence Strategy', Washington, DC, US.

Egel, D, Shatz, H, Kumar, K & Harshberger, ER 2020, *Defense budget implications of the COVID-19 pandemic*, RAND Corporation, viewed on 30 October 2020, <https://www.rand.org/blog/2020/04/defense-budget-implications-of-the-covid-19-pandemic.html>.

Fountaine, T, McCarthy, B & Saleh, T 2019, *What it really takes to scale Artificial Intelligence*, McKinsey Digital, New York, NY, US.

Government Accountability Office (GAO) 2017, 'Military acquisitions, DOD is taking steps to address challenges faced', Washington, DC, US.

Hurley, J & Potter, D 2020, 'Avoiding the pitfalls of an artificial reality', *Proceedings of the 15th International Conference on Cyber Warfare and Security*, Norfolk, VA, US, pp. 236-46.

Lipton, ZC 2020, *The hard problems AI can't (yet) touch*, KD Nuggets, viewed on 30 October 2020, <https://www.kdnuggets.com/2016/07/hard-problems-ai-cant-yet-touch.html>.

Lyle, D 2019, *Overcoming 'the Silo Effect' in the Department of Defense*, NSI, Inc., 5 August, viewed on 30 October 2020, <https://nsiteam.com/social/wp-content/uploads/2019/08/Lyle-IP-2Aug19-Final2R.pdf>.

Marchand, DA, Kettinger, W & Rollins, JD 2000, 'Information orientation: People, technology and the bottom line', *MIT Sloan Management Review*, vol. 42, no. 3, pp. 69-80.

Masunaga, S 2018, 'Google's retreat from AI contract is unlikely to cool the Pentagon's love for Silicon Valley', *Los Angeles Times*, 14 June, viewed on 30 October 2020, <https://www.latimes.com/business/la-fi-dod-silicon-valley-20180614-story.html>.

Morgan, F, Boudreaux, B, Lohn, A, Ashby, M, Curriden, C, Klima, K & Grossman, D 2018, *Military applications of Artificial Intelligence: Ethical Concerns in an uncertain world*, The RAND Corporation, Santa Monica, CA, US.

National Security Commission on Artificial Intelligence 2019, *The Interim Report: November 2019*, viewed 30 October 2019, <https://science.house.gov/imo/media/doc/Schmidt%20Testimony%20Attachment.pdf>.

The National Security Commission on Artificial Intelligence (NSCAI) 2020, *NSCAI First Quarter Recommendations 2020*, NSCAI, Washington, DC, US.

Office of Science and Technology Policy 2019, The White House Summit on Artificial Intelligence in Government 2019, Washington, DC, US.

Polonski, V 2018, *People don't trust AI--Here's how we can change that, Scientific American*, 10 January, viewed 30 October 2020, <https://www.scientificamerican.com/article/people-dont-trust-ai-heres-how-we-can-change-that/>.

Preston, S 2015, *Department of Defense Law of War Manual*, Department of Defense, Washington, DC, US.

Reinsel, D, Gantz, J & Rydning, J 2018, *The Digitization of the World – From Edge to Core*, Doc# US44413318, IDC white paper, November, viewed 30 October 2020, <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>.

Saaty, T 2012, *Decision making for leaders*, RWS, Pittsburgh, PA, US.

Sayler, K & Harris, L 2020, 'Deep fakes and national security', Congressional Research Service (CRS), *CRS in Focus*, 26 August, viewed 30 October 2020, <https://crsreports.congress.gov/product/pdf/IF/IF11333>.

Shanahan, P 2018, Establishment of the Joint Artificial Intelligence Center, Government Executive, viewed on 30 October 2020, <https://admin.govexec.com/media/>.

Taulli, T 2020, 'AI (Artificial Intelligence) governance: How to get it right', *Forbes*, 10 October, viewed on 30 October 2020, <https://www.forbes.com/sites/tomtaulli/2020/10/10/ai-artificial-intelligence-governance-how-to-get-it-right/#10711ef8745f>.

Theohary, C 2018, I*nformation warfare: Issues for Congress*, Congressional Research Service (CRS), Washington, DC, US.

Whaley, D 2019, *The Big Data battlefield, military embedded systems*, 9 August, viewed on 30 October 2020, <https://militaryembedded.com/ai/big-data/the-big-data-battlefield>.

The White House 2018, *National Cyber Strategy of the United States*, Washington, DC, US.

The White House 2019, 'Artificial Intelligence for the American People', Washington, DC, US.

Wirtz, B, Weyerer, J & Geyer, C 2019, 'Artificial Intelligence and the public sector—Applications and challenges', *The International Journal of Public Administration*, pp. 596-615.