2016

8th International Conference on Cyber Conflict:

Cyber Power

N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)

CCDCOE
NATO Cooperative Cyber Defence Centre of Excellence Tallinn Estonia

IEEE
Advancing Technology for Humanity

31 MAY - 03 JUNE 2016, TALLINN, ESTONIA

2016 8TH INTERNATIONAL CONFERENCE ON CYBER CONFLICT: CYBER POWER

IEEE Catalog Number: CFP1626N-PRT
ISBN (print): 978-9949-9544-8-3
ISBN (pdf): 978-9949-9544-9-0

COPYRIGHT AND REPRINT PERMISSIONS

PRINTED COPIES OF THIS PUBLICATION ARE AVAILABLE FROM:

NATO CCD COE Publications
Filtri tee 12, 10132 Tallinn, Estonia
Phone: +372 717 6800
Fax: +372 717 6308
E-mail: publications@ccdcoe.org
Web: www.ccdcoe.org
Head of publishing: Jaanika Rannu
Layout: Jaakko Matsalu

NATO COOPERATIVE CYBER DEFENCE CENTRE OF EXCELLENCE

The Tallinn-based NATO Cooperative Cyber Defence Centre of Excellence (NATO CCD COE) is a NATO-accredited knowledge hub, think-tank and training facility. The international military organisation focuses on interdisciplinary applied research and development, as well as consultations, trainings and exercises in the field of cyber security. The Centre's mission is to enhance capability, cooperation and information-sharing between NATO, Allies and partners in cyber defence.

The heart of the NATO Cooperative Cyber Defence Centre of Excellence is a diverse group of international experts on cyber security. They include legal scholars, policy and strategy experts as well as technology researchers with military, government and industry backgrounds.

Membership of the Centre is open to all Allies. Currently, the Czech Republic, Estonia, France, Germany, Greece, Hungary, Italy, Latvia, Lithuania, the Netherlands, Poland, Slovakia, Spain, Turkey, the United Kingdom and the United States have signed on as Sponsoring Nations of the Centre. Austria and Finland have become Contributing Participants – the status available for non-NATO nations. Furthermore, Sweden and Belgium have announced their intent to join the Centre and the respective negotiations are ongoing.

The Centre is staffed and financed by its sponsoring nations and contributing participants. The Centre is not part of NATO command or force structure, nor is it funded from the NATO budget. Many of the Centre's publications, as well as a number of interactive databases, are accessible through www.ccdcoe.org.

CYCON 2016 SPONSORS

TECHNICAL SPONSOR

DIAMOND SPONSOR

GOLD SPONSORS

SPONSORS

FOREWORD

This is the eighth time that the annual International Conference on Cyber Conflict (CyCon 2016), organised by the NATO Cooperative Cyber Defence Centre of Excellence, is held in the historic city of Tallinn, the capital of Estonia. Over the years the CyCon conferences have proved to be world-recognised forums addressing cyber conflict and security and their implications for society, business and world affairs.

Every year the CyCon conference focusses on a specific aspect of cyber conflict. In 2013, the conference discussed the roles and methods of automated cyber defence and in 2014, it concentrated on active cyber defence. The focus of CyCon 2015 was on architectural aspects of cyberspace. CyCon 2016 concentrates its attention on 'Cyber Power', which can be viewed from very different angles. One might look at it as the ability to exert one's will over the adversary via cyber operations while maintaining the security of one's own systems; or through the perspective of controlling the development and production of software and hardware; or as the issue of trying to make oneself impervious to cyberattack by severely limiting one's dependency on information technology. The possible viewpoints are as different as the actors that discuss them. As is true for all CyCon conferences, the notion of 'Cyber Power' is examined from information technology, strategic, legal, and policy perspectives, in order to provide a comprehensive and well-informed view.

We would like to thank the members of the CyCon 2016 Academic Review Committee and the distinguished peer reviewers for their tireless work in identifying papers for presentation at the conference and for publication in this book. Last, but not least, we are delighted to congratulate the dedicated editors of this volume.

Dr Gabriel Jakobson
Chief Scientist
CyberGem Consulting
USA

Dr Rain Ottis
Associate Professor
Tallinn University of Technology
Estonia

Brookline, Tallinn, April 2016

TABLE OF CONTENTS

INTRODUCTION

For the eighth consecutive year, the NATO Cooperative Cyber Defence Centre of Excellence (NATO CCD COE) is proud to host the International Conference on Cyber Conflict (CyCon), which gathers experts from the government, military, academia and private sector to discuss the most pressing issues related to cyber defence. Once again CyCon serves as an interdisciplinary platform for networking and sharing knowledge, bringing together the cyber security community's technical experts, strategic thinkers, political scientists and lawyers.

CyCon 2016 focuses on understanding the many aspects of 'cyber power' and aims to analyse how new digital technologies are affecting the traditional social, political, and technical foundations of defence and security. Power can be seen as the capacity, ability and willingness to act and is a central element of almost every form of social interaction, be it on the individual or the international level. The rapid development of the digital domain has had a substantial impact on all aspects of society but many questions about how actors can exert power through cyberspace remain unanswered. For instance, how has the rapid rise of information technologies changed the way in which different actors project their influence? How can the strategic and technical aspects of cyber power be measured? Who are the actors seeking cyber power? How do organisations achieve effective defences in cyberspace? What technical, political and legal frameworks are necessary in regard of the build-up of cyber capabilities? How will cyber power impact future military operations?

These and many other questions will be addressed during the conference's sessions and workshops, as well as in the papers that have been selected for the conference proceedings.

This publication, comprising 18 articles, spans a wide spectrum of topics. It starts by focusing on the conceptual issues and asks how the phenomenon of cyber power can be defined and assessed. This is first done by **Jelle van Haaster** who, based on a comparative analysis of traditional approaches to power in international relations, presents a comprehensive methodology for measuring the different dimensions of cyber power. The nature of computer network attacks (CNAs) is then analysed by **Ragnhild Endresen Siedler** as she explores the difference between CNA as a means of pure destruction and CNA as a means of forcible accomplishment. In the following article, **Jason Healey** uses case studies to re-shape our understanding of the often overlooked long-term strategic implications of known cyber operations. The section ends with an article by **Lior Tabansky**, who uses the case of Israel to conceptualise a framework for cyber power on a national level.

Moving away from the more conceptual questions related to cyber power, **Sean T. Lawson et al**. analyse, based on research on fear-appeals, how the public discourse on cyber-doom scenarios affects the development of appropriate cyber security policy measures. Next, **Steve S. Sin et al**. look at the issue of cyber terrorism by estimating the possible cyber capabilities of extremist organisations. The discussion then turns to the 'soft' elements of cyber power. **Drew Herrick** addresses the link between social media and military cyber operations and identifies this connection as a variable for analysing an actor's cyber capabilities. **Pascal Brangetto**

and **Matthijs A. Veenendaal** analyse the growing role of cyber operations as an element of influence operations and introduce the term 'Influence Cyber Operations'.

In the context of the global drive to develop cyber capabilities, the next section of the book focuses mainly on how to limit cyber power, discussing issues of international law and the different options and problems related to developing cyber arms control regimes. **Kubo Mačák** starts by arguing that there is a power vacuum resulting from the reluctance of states to engage in international law-making in the context of cyber security. A theoretical basis for the concept of a cyber arms race dynamic is provided by **Anthony Craig** and **Brandon Valeriano** who examine the rapid cyber weapons build-up by comparing the relations between US-Iran and North Korea-South Korea. **Markus Maybaum** and **Jens Tölle** then look at the lessons that can be learned from the traditional arms control regimes and propose a unique technical solution to bypass the usual issues that are associated with limiting the proliferation of cyber weapons. Possible counter-proliferation methods are further proposed by **Trey Herr** who, in analysing the nature of the malware market and cyber security research, questions the value of the Wassenaar Arrangement.

Moving on to more specific areas of cyber security and defence, **Robert Koch** and **Mario Golling** focus on the cyber security risks associated with the integration of outdated and state of the art military weapon systems. **Kim Hartmann** and **Keir Giles** address a new 'domain' of cyber power as they highlight the growing relevance of unmanned aerial vehicles and present many possible ways in which these can be used in different conflict scenarios. Aviation security is also tackled by **Martin Strohmeier et al**., as they define and discuss a threat model describing the up-to-date capabilities of different types of threat agents and their impact on the new digitalised aviation communication systems.

The book ends with articles that aim to provide practical solutions to several specific cyber security issues related to privacy and advanced persistent threats (APTs). First, **Mirco Marchetti et al**. focus on mitigating the threat from APTs by proposing a framework that incorporates techniques based on big data analytics and security intelligence. Second, **Paolo Palmieri** analyses the protection of anonymity networks as he presents the structural weaknesses of Tor and proposes a number of modifications to improve its resilience against DDoS attacks. Finally, **Siddharth Prakash Rao et al**. draw attention to the vulnerabilities in the existing cellular networks used for location tracking and surveillance and highlight the existing countermeasures to solve the issue.

All of the articles in this book have been through a double-blind peer review by the Academic Review Committee. We would therefore like to thank the Co-Chairs as well as the distinguished members of the Academic Review Committee for their efforts in reviewing, discussing and selecting the submitted papers, guaranteeing their academic quality.

Academic Review Committee Members:

- Prof Gabriel Jakobson, CyberGem Consulting, Co-Chair of the Academic Review Committee

- Dr Rain Ottis, Tallinn University of Technology, Co-Chair of the Academic Review Committee
- Dr Iosif Androulidakis, Ioannina University
- Bernhards Blumbergs, NATO CCD COE
- Maj Pascal Brangetto, NATO CCD COE
- Dr Russell Buchan, University of Sheffield
- Dr Steve Chan, MIT; Harvard University
- Prof Thomas Chen, Swansea University
- Prof Michele Colajanni, University of Modena and Reggio Emilia
- Dr Christian Czosseck, NATO CCD COE Ambassador
- Prof Dorothy E. Denning, Naval Postgraduate School
- Prof Gabi Dreo Rodosek, Bundeswehr University
- BGen Prof Paul Ducheine, Amsterdam University
- Dr Kenneth Geers, NATO CCD COE Ambassador
- Keir Giles, Conflict Studies Research Centre
- Prof Michael Grimaila, Air Force Institute of Technology
- Prof Dimitris Gritzalis, University of Economics of Athens
- Dr Jonas Hallberg, Swedish Defence Research Agency (FOI)
- Jason Healey, Columbia University
- Margarita Levin Jaitner, Swedish Defence University
- Maj Harry Kantola, NATO CCD COE
- Kadri Kaska, NATO CCD COE
- Prof Sokratis Katsikas, University of Piraeus
- Mari Kert-St Aubyn, NATO CCD COE
- Prof Jörg Keller, Hagen Open University
- Prof Panagiotis Kikiras, AGT International
- Dr Marieke Klaver, TNO
- Prof Konstantinos Lambrinoudakis, University of Piraeus
- Dr Scott Lathrop, United States Military Academy
- Dr Sean Lawson, University of Utah
- Corrado Leita, LASTLINE Inc
- Prof Jarno Limnéll, Aalto University
- Dr Lauri Lindström, NATO CCD COE
- Eric Luiijf, TNO
- Dr Matti Mantere, Intel
- Cpt Markus Maybaum, NATO CCD COE Ambassador
- Prof Michael Meier, Bonn University
- Tomáš Minárik, NATO CCD COE
- Dr Jose Nazario, INVINCEA Inc
- Dr Lars Nicander, Swedish National Defence College
- Anna-Maria Osula, NATO CCD COE
- Liisa Past, NATO CCD COE
- Dr Patryk Pawlak, EU Institute for Security Studies
- Raimo Peterson, NATO CCD COE

- Mauno Pihelgas, NATO CCD COE
- Maj Nikolaos Pissanidis, NATO CCD COE
- Lt-Col Jari Rantapelkonen, Finnish National Defence University
- Henry Rõigas, NATO CCD COE
- Prof Juha Röning, University of Oulu
- Julie J.C.H. Ryan, George Washington University
- Lt-Col Jan Stinissen, The Netherlands Ministry of Defence
- Lorena Trinberg, NATO CCD COE
- Dr Lauri Tuovinen, University of Oulu
- Dr Jens Tölle, Fraunhofer FKIE
- Dr Enn Tõugu, Tallinn University of Technology
- Dr Risto Vaarandi, Tallinn University of Technology
- Teemu Uolevi Väisänen, NATO CCD COE
- Lt-Col Jens van Laak, NATO CCD COE
- Matthijs Veenendaal, NATO CCD COE
- Prof Ari Visa, Tampere University of Technology
- Dr Jozef Vyskoc, VaF Rovinka and Comenius University Bratislava
- Prof Bruce Watson, Stellenbosch University
- Dr Sean Watts, Creighton University
- Prof Stefano Zanero, Milan University

Special thanks are due to the Institute of Electrical and Electronic Engineers (IEEE) as the IEEE Estonia Section served as a technical co-sponsor for CyCon 2016 and this publication.

We would also like to express our gratitude to Jaanika Rannu and others in the NATO CCD COE supporting staff for their excellent organising skills and assistance during the publication process.

Last but not least, we would like to thank all the authors of the papers collated in this publication for their excellent submissions, friendly cooperation and their commitment to expanding the scope of cyber security studies.

The CyCon 2016 Agenda Management Board

Maj Pascal Brangetto
Cpt Raik Jakschis
Mari Kert-St Aubyn
Lauri Lindström
Maj Nikolaos Pissanidis
Henry Rõigas
Teemu Uolevi Väisänen
Matthijs Veenendaal

NATO Cooperative Cyber Defence Centre of Excellence
Tallinn, Estonia, April 2016

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Assessing Cyber Power

Jelle van Haaster
Faculty of Military Sciences
Netherlands Defence Academy, Breda
University of Amsterdam
j.vanhaaster@uva.nl

Abstract: This paper aims to contribute to the debate regarding quantitative and qualitative appreciation of cyber power. It will do so by: (1) deriving a thorough foundation for the cyber power discussion from the 20th century power debate; (2) presenting a comprehensive framework for analysing (cyber) power; and (3) positioning cyber capacities within this framework. Revamping the 20th century power discussion is necessary as the current cyber power debate focuses much on the 'means' component of power (e.g. DDoS capacity, network infrastructure spending, malware acquisition budgets, etc.). This view of power is very similar to the pre-World War II approach to assessing power. The power theorists, however, have shied away from this approach as it proved to be too narrow. Primarily because it failed to capture why a more resourceful actor sometimes fails to ascertain its objectives or preferred outcomes vis-à-vis a smaller actor (e.g. the United States' experience in Vietnam). In order to fill this lacuna, this paper offers a more comprehensive approach to power, including all dimensions of (cyber) power, being: scope, domain, weight, costs and means.

Keywords: *power, cyber power, quantification, cyber arms race, international relations*

1. INTRODUCTION

States relish comparing their own power with that of other states. When it stacks up, there is no need to worry, but whenever it appears to fall short it is deemed an omission requiring attention. Seemingly empiric quantification of power drives this type of governmental decision-making. By quantifying power, the feasibility and probable efficacy of a particular action can be determined. Statements about feasibility and efficacy are based on the notion that more powerful states, in whatever respect, are more likely to be able to advance their goals than weaker states.

Ranking of states on whatever basis frequently occurs, either by states or in the media. These forms of categorisation, qualitative appreciation, and quantification of power and power

resources at a state's disposal received considerable academic attention throughout the twentieth century. Although power and its quantification remain unsettled issues within the study of international relations, the contours of the discussion have settled.

Faced with new possibilities, challenges, and risks of interconnection on unprecedented scale, new questions arise with regard to the use and quantification of power, particularly questions about conveying power in or through cyberspace. Contemporary scholarly research expresses that cyberspace serves as a potent conduit through which power can be conveyed. It has not, however, addressed issues with regard to feasibility and efficacy of a particular course of action, or the preceding assessment of the distribution of cyber power. The contemporary cyber power debate seemingly also omits the power debate from the 20th century. Instead of using the conclusions of this debate, many use a pre-World War II approach to power, namely, that power is epitomised in control over resources. In order to fill this lacuna this paper will revamp the 20th century power discussion and adjoin it with contemporary insights regarding cyber power and capacities.

This paper will first briefly describe the essential elements of the twentieth century power discussion regarding the quantitative and qualitative appreciation of power. Section 2 will conclude with a conceptual framework for analysing power. After that, the paper will highlight the contemporary cyber power discussion in Section 3. In doing so, this paper will discuss various viewpoints on cyber power's quantitative and qualitative assessment. In Section 4, the framework for analysing power developed in Section 2 will be adjoined with contemporary notions regarding cyber capacities. In doing so, this paper will provide: (a) a basis for further discussion about cyber power not omitting more than a hundred years of power debate; (b) footholds for analysing cyber power; and (c) a viewpoint on the categorisation of cyber capacities within the power debate.

2. POWER

Within international relations, the concept of power is one of the most contested issues.[1] Even though 'weighty books have analysed and elaborated the concept',[2] much is unsettled apart from a notion that the issue requires attention.[3] It especially requires attention in the context of this paper, which seeks to present how cyber power can be assessed.

Comparing one's power to another has been a core activity for rulers from antiquity until now. Describing power as a concept and determining how to measure it is a more 'recent' development beginning in the late 18th century. Although seen from an entirely different *zeitgeist*, the mid- and late-20th century power debate has yielded valuable insights. This section will first briefly discuss different viewpoints on power derived from that discussion and then examine perspectives on assessing power.

This section will only cover Barnett and Duvall's categorisation of power's operation as it captures the most important viewpoints of the 20th century power discussion. Their insights

1 David A. Baldwin, 'Power and International Relations,' in *Handbook of International Relations*, eds. Walter Carlsnaes, Thomas Risse and Beth A. Simmons (London: Sage, 2002), 177-191. p.177.
2 Robert Gilpin, *War and Change in World Politics* (Cambridge: Cambridge University Press, 1983). p.13.
3 Ronen Palan, *Global Political Economy: Contemporary Theories* (London: Routledge, 2000). pp.53-54.

are based upon many fundamental post-World War II viewpoints on power.[4] This paper will not touch upon the pre-World War II viewpoints on power other than to observe that the notion at the heart of this 'elements of national power' approach was that a specific amount of power in the form of resources can be possessed, that the actor with most resources is most likely to forward their interests, and hence is more powerful.[5]

2.1 The operation of power

Where and how does power manifest itself? Debate regarding this question has been going on from the 1960s in the form of the faces of power (or power dimensions) debate, and has quite recently resulted in the more modern 'taxonomy of power concepts framework' described by Barnett and Duvall.[6]

Barnett and Duvall distinguish four types of power:

- Compulsory, epitomising 'power as relations of interaction of direct control by one actor over another';
- Institutional, considering 'the control actors exercise indirectly over others through diffuse relations of interaction';
- Structural, expressing 'the constitution of subjects' capacities in direct structural relation to one another'; and
- Productive, entailing the 'socially diffuse production of subjectivity in systems of meaning and signification'.[7]

This subsection will briefly describe these four power concepts.

Compulsory power follows the Dahlian definition of power: 'the ability of A to get B to do what B otherwise would not have done'.[8] It is very similar to the first face of power, and hinges on the intentionality, conflict, and successfulness of A.[9] Both state and non-state actors can exert compulsory power, 'multinational corporations can use their control over capital to shape the foreign [and global] economies' and 'non-state networks and groups sometimes […] terrorise entire populations'.[10] Compulsory power does not require material resources: 'it also entails symbolic and normative resources'.[11]

4 See for instance: Robert A. Dahl, 'The Concept of Power,' *Behavioral Science* 2, no. 3 (1957), 201-215.; Robert A. Dahl, 'A Critique of the Ruling Elite Model,' *The American Political Science Review* 52, no. 2 (1958), 463-469.; Klaus Knorr, *The Power of Nations: The Political Economy of International Relations* (New York: Basic Books, 1975).; Harold Hance Sprout and Margaret Tuttle Sprout, *Foundations of International Politics* (New Jersey: Van Nostrand, 1962).

5 See for instance: Hans J. Morgenthau, *Politics among Nations: The Struggle for Power and Peace* (New York: Alfred A. Knopf, 1948).; Hans J. Morgenthau, *Dilemmas of Politics* (Chicago: University of Chicago Press, 1958).; Quincy Wright, *A Study of War: Volume II* (Chicago: The University of Chicago Press, 1942).; J. D. Singer and Melvin Small, 'The Composition and Status Ordering of the International System: 1815-1940,' *World Politics* 18, no. 2 (1966), 236-282.; Frank H. Simonds and Brooks Emeny, *The Great Powers in World Politics: International Relations and Economic Nationalism* (New York: American Book Company, 1937).; A. F. Pollard, 'The Balance of Power,' *Journal of the British Institute of International Affairs* 2, no. 2 (1923), 51-64.

6 Michael Barnett and Raymond Duvall, 'Power in International Politics,' *International Organisation* 59, no. 1 (2005), 39-75.

7 Ibid. p.43.

8 Dahl, *The Concept of Power*. pp.202-203.pp.202-203.

9 Barnett and Duvall, *Power in International Politics*. p.49.

10 Ibid. p.50.

11 Ibid. p.50.

Institutional power involves an:

> 'actors' control of others in indirect ways […] the conceptual focus here is on the formal and informal institutions that mediate between A and B, as A, working through the rules and procedures that define those institutions, guides, steers, and constraints the action (or non-actions) and conditions of existence of others'.[12]

In such a scenario, A does not exercise power directly over B, 'A cannot necessarily be said to 'possess' the institutions that constraints and shapes B', but A could be the dominant actor 'that maintains total control over an institution'.[13]

Structural power 'concerns the structures – or, more precisely, the co-constitutive internal relations of structural positions – that define what kinds of social beings actors are'.[14] It 'concerns the determination of social capacities and interests' of actors, based upon the notion 'that the structural position of A exists only by virtue of its relation to the structural position of B'.[15] It 'is the production and reproduction of internally related positions of super- and subordination, or domination, that actors occupy'.[16] Structural power is best characterised by Steven Lukes' statement that it is 'the supreme and most insidious exercise of power' to prevent or permit actors from arising within societies or structures.[17]

Productive power is based on more or less 'generalised and diffuse social processes', unlike structural power that is based on direct structural relations.[18] Productive power 'is the constitution of all social subjects with various social powers through systems of knowledge and discursive practices of broad and general scope'.[19] In other words, productive power looks beyond structures, it is concerned with 'the social processes and the systems of knowledge through which meaning is produced, fixed, lived, experienced and transformed',[20] but also how discursive processes and practices produce social identities and capacities'.[21] Examples of productive power is 'the discursive production of subjects by using categories of classification such as 'civilised, rogue, European, unstable, Western, and democratic states'.[22]

2.2 Assessing power

The power of states was deemed to be easily measurable in the eighteenth century.[23] Factors taken into account were 'territory, wealth, armies and navies'.[24] The eighteenth century concept of quantifiable power based on resources has played a prominent role throughout history. Although some additions have been made, many decision-makers and political theorists still

12 Ibid. p.51.
13 Ibid. p.51.
14 Ibid. pp.52-53.
15 Ibid. p.53.
16 Ibid. p.55.
17 Peter Digeser, 'The Fourth Face of Power,' *The Journal of Politics* 54, no. 4 (1992), 977-1007. p.979.; Barnett and Duvall, *Power in International Politics*, 39-75. p.53.
18 Ibid. p.5.
19 Ibid. p.55.
20 Ibid. p.55.
21 Ibid. p.56.
22 Ibid. p.56.
23 *Baldwin, Power and International Relations*. pp.177-178.
24 Ibid.

believe that a state's power position can be derived from its sources of power, and that power can be possessed, stored, and collected.[25]

The 'relational power approach' conceives power and its measurability differently from the 'elements of national power approach'. Proponents of this approach argue that power should be defined relationally: it 'does not reside in a resource but stems from the particular relation in which abilities are actualised'.[26] Power is only meaningful 'if its control is seen to be valued by other actors in the interaction'.[27] Within this approach power is deemed multidimensional[28] and dependent of 'specific policy-contingency frameworks'.[29] This subsection will briefly describe these aspects.

Power is multidimensional. It consists – at least – of the dimensions of scope and domain.[30] Less accepted domains are dimensions such as weight, costs, and means.[31] The scope of power is understood to comprise objectives and the affected issue areas.[32] Domain 'refers to the […] actors subject to [the] influence [attempt]'[33] or simply 'the target[s]'.[34] Weight relates to the potential effectiveness of power; that is, the likelihood that 'B's behaviour is or could be affected by A'.[35] Costs indicate both the cost to actor A and the costs to B; for instance 'is it costly or cheap for A to influence B? Is it costly or cheap for B to comply with A's demands?'[36] Means refer to the various instruments 'of exercising influence', there are 'many ways to categorise such means', and these various instruments will be discussed in section four.[37]

Any statement about power would be meaningless without 'the specification of the situation' or the context.[38] Specifying the context is 'the single most important step in capability analysis' and basically entails 'establishing who is trying to get whom to do what' and in what situation.[39] Some power resources may be useless in one situation, whilst being extremely influential in others, 'the only way to determine whether something is a power resource or not is to place it

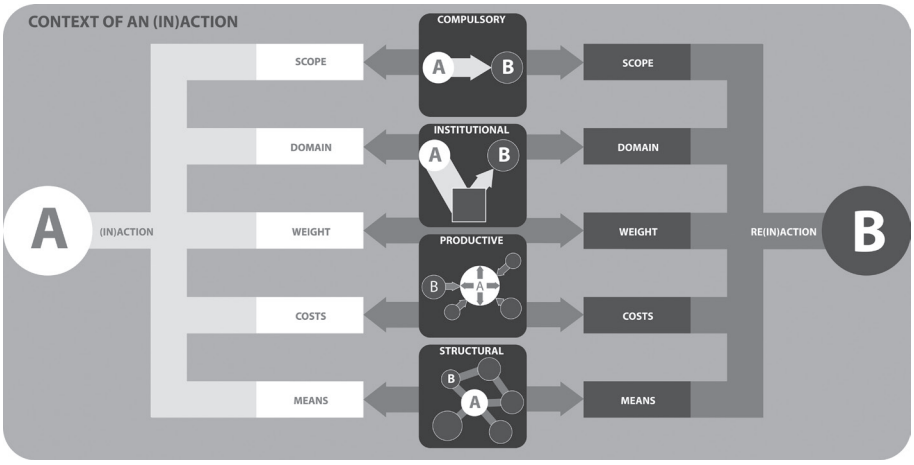
25 See for instance: Morgenthau, *Politics among Nations: The Struggle for Power and Peace.*
26 Stefano Guzzini, 'On the Measure of Power and the Power of Measure in International Relations,' *DIIS Working Paper*, no. 28 (2009). p.7.
27 Stefano Guzzini, 'Structural Power: The Limits of Neorealist Power Analysis,' *International Organisation* 47, no. 3 (1993), 443-478. pp.452-453.
28 Baldwin, *Power and International Relations*. p.178.
29 Guzzini, *Structural Power*. p.453.
30 Harold D. Laswell and Abraham Kaplan, *Power and Society: A Framework for Political Inquiry* (New Haven: Yale University Press, 1950). p.74.; Baldwin, *Power and International Relations*. p.179.; Joseph S. Nye, *The Future of Power*, 1st ed. (New York: Public Affairs, 2011). p.6.; Dahl, *The Concept of Power*. p.203.
31 Baldwin, *Power and International Relations*. p.178.; See also: Laswell and Kaplan, *Power and Society: A Framework for Political Inquiry*. p.74. They describe the dimensions domain, scope, weight and coerciviness; Dahl, *The Concept of Power*. p.203. He describes the dimensions of base/domain, means/instruments, amount/extent and range/scope.
32 Nye, *The Future of Power*. p.6.; Baldwin, *Power and International Relations*. p.180.; Guzzini, *Structural Power*. p.453.
33 Baldwin, *Power and International Relations*. p.180.
34 Guzzini, *Structural Power*. p.453.
35 Baldwin, *Power and International Relations*. p.180.; Guzzini, *Structural Power: The Limits of Neorealist Power Analysis*, 443-478. pp.453-454.; See also: Dahl, *The Concept of Power*. p.203. He refers to weight simply as the amount or extent an actor has power over another actor.
36 Baldwin, *Power and International Relations*. p.178.
37 Ibid. pp.178-179.
38 Guzzini, *Structural Power*. p.454.
39 David A. Baldwin, *Economic statecraft* (New Jersey: Princeton University Press, 1985). p.149.

in the context of a real or hypothetical situation.[40] In order to determine the context, amongst other, the historical and societal background should be analysed.[41]

2.3 Summary

Although the theoretical foundations of the 'elements of national power approach' and relational power differ with regard to the possession of power and its measurability, the approaches are not irreconcilable. The point of the relational power approach is that a focus on a single or particular set of dimensions could result in tunnel vision, potentially ruling out equally important dimensions. The relational power approach includes the 'means' dimension forwarded by the 'elements of national power approach' proponents, but it adjoins it with other dimensions. As such, the relational power approach is comprehensive. This section will conclude with a framework for analysing power integrating the approaches (see figure 1).

FIGURE 1: POWER ANALYSIS FRAMEWORK. Power manifests itself in a relation between actors; actor A has power to the extent that actor B is receptive to that particular form of power. Whether or not B is receptive depends on a variety of factors, first and foremost the context. Other factors to be taken into account when determining receptiveness are the scope (what is the objective), domain (what actors are involved), weight (likelihood of effectiveness), costs (how costly is the action for A and for B to comply) and means or instruments involved. Barnett and Duvall's concepts of power taxonomy serves to illuminate the power arena of actor's (in)actions and re(in)actions. The power concept utilised in A's (in)action also influences receptiveness and successfulness (compulsory, institutional, structural and/or productive).



3. ASSESSING CYBER POWER

After having discussed the 'old' notion of power, this paper will now reflect on cyber power and its assessment. First, it will define the etymological approach taken to cyber power in this paper, and then sketch the contours of the current cyber power debate by looking at the work of Nye and of Betz and Stevens.

40 David A. Baldwin, 'Power Analysis and World Politics: New Trends Versus Old Tendencies,' *World Politics* 31, no. 2 (1979), 161-194. p.165.

41 Guzzini, *Structural Power*. p.454.

3.1 Cyber power

The meaning of the cyber prefix has undergone some radical changes over time, both in etymological and political senses. Etymologically, it went from being a prefix pertaining to the government element in cybernetics,[42] to a word describing the ethereal world of data,[43] to the contemporary notion in which cyber and cyberspace can be used interchangeably with the Internet, networks and computers. From a security point of view, it went from computer security in the 1960s and 1970s,[44] to information security in the 1980s, resulting in today's coexistence of both information- and cyber-security. From a political viewpoint, cyber went from being a computer threat to critical or vital infrastructure[45] to being both vulnerability and an opportunity to be exploited by intelligence agencies and the military.[46] As a concept of military doctrine it went from being an overarching war fighting concept highlighting the prime role of information on the battlefield,[47] to a current divide over whether or not cyber operations are a subset of information operations.[48]

This paper will take the approach to 'cyber' as pertaining to the intrinsic character of the means and methods used. The specific character lies in its origin, i.e. cyberspace and its specific destination, being other entities within cyberspace. As a result, cyber power would entail conveying power in or through cyberspace. An assessment of cyber power consequently would involve an estimation of an actor's ability to convey power in cyberspace. It can best be understood as: 'the variety of powers that circulate in cyberspace and which shape the experiences of those who act in and through cyberspace'.[49]

Cyberspace refers to the construct created by governmental decision- and policymakers. This is the environment 'formed by physical and non-physical components, characterised by the use of computers and the electro-magnetic spectrum, to store, modify and exchange data using computer networks'.[50] There are more detailed conceptualisations of this domain, which describe various components in cyberspace, namely: geographic, which involves the location; physical, which comprises the network hardware and infrastructure; logical, which captures the software and logical connections; and cyber persona online profiles (such as social-media profiles and mail accounts).[51]

42 See for instance: Norbert Wiener, *The Cybernetics of Society: The Governance of Self and Civilisation* (Cambridge: M.I.T. Press, 1948).

43 See for example: William Gibson, *Neuromancer* (New York: Berkley Publishing Group, 1984).

44 See for instance: Michael Warner, 'Cybersecurity: A Pre-History,' *Intelligence and National Security* 27, no. 5 (2012), 781-799. pp.787.

45 See for instance: The White House, *Securing America's Cyberspace, National Plan for Information Systems Protection: An Invitation to a Dialogue* (Washington, DC: The White House, 2000).

46 See for instance: The Joint Chiefs of Staff, *The National Military Strategy of the United States of America: A Strategy for Today; A Vision for Tomorrow* (Washington, DC: Office of the Chairman, 2004). p.18; The Chairman of The Joint Chiefs Of Staff, *The National Military Strategy for Cyberspace Operations* (Washington, D.C.: Office of the Chairman, 2006). p.3.

47 See for example: John Arquilla and David Ronfeldt, *Networks and Netwars: The Future of Terror, Crime, and Militancy* (Santa Monica: RAND National Defense Research Institute, 2001).

48 See for instance: North Atlantic Treaty Organisation, *Allied Joint Publication 3.10: Allied Joint Doctrine for Information Operations* (Brussels: North Atlantic Treaty Organisation, 2009). pp.1-7 to 1-13.; The Joint Chiefs Of Staff, *Joint Publication 3-12 (R): Cyberspace Operations* (Washington, D.C.: The Joint Chiefs Of Staff, 2013). pp.5-6.

49 Betz and Stevens, *Cyberspace and the State: Toward a Strategy for Cyber-Power.*

50 Michael N. Schmitt, *Tallinn Manual on the International Law Applicable to Cyber Warfare* (Cambridge: Cambridge University Press, 2013). p.258.

51 See for instance: The Joint Chiefs Of Staff, *Joint Publication 3-12 (R): Cyberspace Operations*. pp.I-2 and I-3; United States Army, *Cyberspace Operations Concept Capability Plan 2016 2028* (Fort Eustis: TRADOC, 2010). p.9.

Combining the general notion of cyber power with the more detailed description of cyberspace results in the following notion of cyber power: cyber power comprises the variety of powers affecting the geographic, physical network, logical, and cyber persona components, which consequently shape the experiences of state and non-state actors who act in and through cyberspace. This includes, for instance, using social-media profiles (the cyber persona component) to affect others; the use of offensive cyber means and methods to digitally compromise a critical system (the logical component); or using law enforcement or military powers to physically establish control over network infrastructure (a physical network component).

The proposed notion of cyber power results in all human-induced activities being qualified as cyber power, including routine uses of cyberspace such as emailing colleagues, texting a friend, or posting a Facebook update. This paper will argue that this is indeed cyber power, albeit a very Foucauldian notion of cyber power. Foucault argued that we are moulded and affected by series of discursive power processes, 'power is co-extensive with the social body; there are no spaces of primal liberty between the meshes of the network'.[52]

Every action we take constitutes or is affected by series of social power processes affecting individual decision-making, such as our psyche, subjectivity, personality, consciousness. These power processes influence us constantly, for instance when consuming information (scrolling through social-media timelines only expressing success), responding or not to mails (expressing a power configuration in which the receiver feels obliged to respond promptly or can afford to not respond), and updating social-media profiles (potentially enforcing one's position in a network).

Although it may constitute cyber power, these processes are virtually impossible to unveil and assess. As such, for practically assessing cyber power the scope of Foucauldian cyber power may be too broad and of little use to decision-makers. Adding a form of intentionality would prove much more practical, resulting in cyber power related to the situations in which an actor is *not unintentionally* trying to improve their power position by using cyberspace components. The following section will adjoin this conceptual view of cyber power with more concrete examples.

3.2 Assessing cyber power

How should cyber power be assessed? Betz and Stevens have applied Barnett and Duvall's taxonomy of power concept to analyse cyber power.[53] A similar approach was taken by Joseph Nye who drew on the faces of power discussion and infused it with his hard and soft power theory to analyse power in cyberspace.[54] As a point of departure for assessing cyber power, Nye and Betz & Stevens are more than useful (see table 1).

52 Michel Foucault, *Discipline and Punish: The Birth of the Prison*, 2nd ed. (New York: Random House, 1995).; See also: Digeser, *The Fourth Face of Power*, 977-1007.

53 Betz and Stevens, *Cyberspace and the State*. pp.x-xx.

54 Joseph S. Nye, *Cyber Power* (Cambridge: Harvard Kennedy School Belfer Center for Science and International Affairs, 2010).

TABLE 1: OVERVIEW OF CYBER CAPACITIES TO BE CONSIDERED WHEN ASSESSING CYBER POWER ACCORDING TO NYE, AND BETZ & STEVENS

Joseph Nye's[55] forms of cyber power	Betz and Stevens'[56] forms of cyber power
First face (A induces B to do what B would otherwise not do) *Hard power* • (Distributed) Denial of service attacks • Insertion of malware • SCADA/ICS disruptions • Arrests of bloggers *Soft power* • Information campaigns **Second face (agenda control)** *Hard power* • Firewalls, filters, and pressure to exclude some ideas *Soft power* • Self-monitoring of ISPs and search engines • ICANN rules on domains • Software standards **Third face (preference shaping)** *Hard power* • Threats to punish bloggers *Soft power* • Information to create preference • Develop norms of revulsion	**Compulsory (direct coercion)** • Control of machines or networks • Deploying non-material resources (e.g. threats) **Institutional (via institutions)** • Influence behaviour through institutions • Set norms and standards • Influence foreign audiences via media institutions **Structural (influencing structures)** • Changing structures (e.g. hierarchical to networked) **Productive (constitution of the subject)** • Reproduce and reinforce existing discourses • Construct and disseminate new discourses

Nye earmarks specific means and methods with an effect in cyberspace, and using these would constitute power. The cyber capacities expressed by Nye are almost universal. Whether explicitly or not, every actor has the ability to execute denial of service attacks, issue threats, repress ideas, and conduct information campaigns. Since virtually every actor has these capacities, its use as an analytical tool for policy- and decision makers is doubtful, although it is a very valuable academic insight. A question begging an answer is: how much? How much denial of service, malware insertion and other cyber capacities does one need? And how could one go about assessing how much is required to be successful? Having defined the relational approach to power as leading in this paper, the only logical answer is: it depends on the context and other dimensions of power.

Betz and Stevens have a different, and more conceptual approach to power. As they use Barnett and Duvall's taxonomy of power, they highlight the different processes for conveying power. It is the 'arena' where the battle for power is fought as opposed Nye's description of the potential 'weapons' with which the battle is conducted. As a matter of analysis, Betz and Steven serve the power debate best by showing that Barnett and Duvall's taxonomy can be applied to cyber power. Again, from a policy- and decision making perspective, it adds little to Nye's categorisation.

As discussed above, means alone do not constitute power, and the same goes for the means described by Nye and Betz & Stevens. Stepping into the pitfall of the single-facet approach to power, for instance by only considering the means, would lead to an arbitrary depiction of the

55 Ibid. p.5.
56 Betz and Stevens, *Cyberspace and the State*. pp.45-53.

power situation. The following section will highlight how to assess cyber power by using the power analysis framework depicted in Figure 1.

4. CYBER POWER CONSIDERATIONS

As mentioned somewhat cynically in the introduction, policy- and decision-makers 'relish' comparing their power to that of other actors. Most often, however, this comparison is a dire necessity in the realm of conflict and when considering the viability of a particular course of action. Hence, the practice of comparing power has its value in political and strategic decision-making.

Unfortunately is seems as if we have forgotten the 20th century power debate and start all over when considering cyber capacities. That debate started with the control over resources approach; the actor surpassing the opponent in access to resources was deemed more powerful. We are currently at this stage in assessing the notion of cyber power. We deem the actor with more resources, be they financial, infrastructural, intellectual, or human, to be the more powerful. Although Nye as well as Betz and Stevens use concepts derived from the post-World War II power discussion, the notion lying at heart of their argument is that power can be possessed; the one with the greatest access to resources and means is the most powerful.

This paper deems power to be relational; power manifests itself in a relationship between actors. Whether or actor A is able to influence actor B depends on the context, scope, domain, weight, cost, and means. The following subsections will discuss these dimensions and link them to cyber capacities.

4.1 Context

The context has implications for the effect of an action; the effectiveness of an action depends strongly on the context. The effect of using cyber capacities in the context of a border dispute differs from using them in a trade conflict.[57] There are myriads of contingencies and contexts in which cyber capacities may or may not prove to be effective. Taking heed of the context is essential for analysing the power situation. Any statement on cyber power should be preceded with describing the context of the power use.

4.2 Scope

The scope of an action is about the effect actor A seeks to achieve in actor B; it is about delineating the objective of a certain action. In inter-state affairs there are various objectives sought after at a strategic level, unfortunately, there is no generalised overview of these objectives. As an illustration of potential objectives, this paper will use the strategic objectives, or strategic functions/effects described by the Dutch government, namely: anticipation, prevention, deterrence, protection, intervention, stabilisation, and normalisation.[58]

Taking the scope into account when assessing cyber power is essential as the effectiveness of a particular cyber capability depends greatly on the objective sought. Some capacities are better

57 Baldwin, P*ower Analysis and World Politics*. p.164.
58 Ministry of Defence (Netherlands), *Future Policy Survey: Summary and Conclusions* (The Hague: MoD, 2010). pp.14-15.

suited for achieving a specific objective than others. For example, a state-owned or bought botnet may serve as a deterrent – that is, if their existence is disclosed – but may have very little effect in stabilising another country. The same goes for disclosing information regarding state-owned, bought, or distributed malware;[59] budget increases in the realm of cyber capacities;[60] or a speech dissuading any potential adversary from impinging on state affairs.[61] These may have deterrent, preventative, or protective effects, but they will lack effectiveness in achieving normalisation or stabilisation of a situation. The use of communication channels on the Internet such as social-media (Twitter, Facebook, YouTube) and other media (instant messaging, 'regular' messaging) may contribute to stability and normalisation in a region,[62] although they may lack a deterrent effect. The scope of the action and its objectives influence whether or not actor A can effectively influence actor B.

4.3 Domain

Domain 'refers to the […] actors subject to [the] influence [attempt]'[63] or simply 'the target[s]'.[64] It is about the type and number of actors influenced by actor A. The domain is crucial to assess power, as the effectiveness of a particular action depends greatly on who actor B is, and how many B's there are. Depending on the domain, or the targets, taking a particular course of action may or may not be effective. Cultural, political and many other aspects of actor B may render it unreceptive to influence exerted by actor A. For instance, a state with strict media controls or censorship in the physical and virtual domain may be unreceptive to inputs from actor A in the realm of social and conventional media. Also, if the domain comprises more than one actor, this will have an impact on the potential effectiveness of actor A's influence attempt. The target actors may differ in preference or location, possibly making it harder for actor A to effectively influence them, and so the nature and number of actor B will greatly influence the effectiveness of a particular course of action undertaken by A.

4.4 Weight

The weight relates to the (potential) effectiveness of power; that is, the likelihood that 'B's behaviour is or could be affected by A'.[65] The actor who has a higher chance of achieving its objectives (weight) can be considered, in a specific context, to be more powerful. The weight of an actor's action depends on all other power dimensions. As such, it may be perceived as the sum of all dimensions and an indicator to who is powerful in a given situation.

4.5 Costs

Costs indicate both the cost to actor A and the costs to B; 'is it costly or cheap for A to influence

59 'Special Report: U.S. Cyberwar Strategy Stokes Fear of Blowback,' Reuters, last modified May 10, accessed December 23, 2015, reuters.com/article/usa-cyberweapons-idINDEE9490AX20130510.

60 See for instance: 'Cyber Command's Exploding Budget,' The Washington Post, last modified January 15, accessed December 23, 2015, washingtonpost.com/news/the-switch/wp/2014/01/15/cyber-commands-exploding-budget-in-1-chart/.

61 See for example: 'U.S. Decides to Retaliate Against China's Hacking,' The New York Times, last modified July 31, accessed December 23, 2015, nytimes.com/2015/08/01/world/asia/us-decides-to-retaliate-against-chinas-hacking.html?_r=0.

62 See for instance, community policing via social-media: 'Kenyan Chief's Twitter Feed Helps Round Up Stolen Cows and Lost Phones,' Los Angeles Times, last modified September 1, accessed December 23, 2015, latimes.com/world/great-reads/la-fg-c1-kenya-twitter-20150901-story.html.

63 Baldwin, *Power and International Relations*. p.180.

64 Guzzini, *Structural Power*. p.453.

65 Baldwin, *Power and International Relations*. p.180; Dahl, *The Concept of Power*. p.203. He refers to weight simply as the amount or extent an actor has power over another actor.

B? Is it costly or cheap for B to comply with A's demands?'[66] These costs for A and B are indicative of power, and 'some have suggested that more power should be attributed to an actor that can exercise influence cheaply'.[67] For example, if it is cheaper for A to influence than for B, actor A is deemed more powerful. Or when A can cheaply influence B to do something costly, A is considered more powerful. In the realm of cyber capacities, costs are an interesting dimension to power. Some capacities are very asymmetric, that is, low cost and high return. For instance a distributed denial of service attack costs very little,[68] and can cause great financial damage[69] by necessitating DDoS mitigation and forensic services, and the financial loss due to the inability to conduct business. Not only a low level attack such as a DDoS is asymmetrical, even a high-end, costly, multi-year intrusion operation like Stuxnet, Flame or Duqu may cause more damage in financial, political, or symbolic senses than its development cost. In other words, costs should be weighted against the benefits or effects of the action.

4.6 Means

Means refer to the various instruments of exercising influence, and there are many ways to categorise such means. Various authors have forwarded categorisations of instruments of state power or statecraft. Military scholarly research often uses the diplomacy, informational, military, and economic (DIME) categorisation, a concept spawned during the Cold War.[70] In the post 9/11 decade, financial, intelligence, law enforcement and 'other civil capacities' instruments were added, resulting in the DIMEFIL or MIDLIFE acronym.[71] There are many other categorisations rivalling or, sometimes, dwarfing the DIMEFIL construct in comprehensiveness and academic stature. This subsection will first briefly discuss the instruments of state power, and then forward an overview of the means enclosed in the DIME, DIMEFIL and other categorisations. After that, these means will be supplemented with cyber capacities described by Nye and Betz & Stevens, and other cyber capacities.

Although the instruments of state power are not subject to much debate, the 'terminology in this domain is not widely agreed upon'.[72] Carr's 1939 *The Twenty-Years' Crisis* serves as the starting point for most writings on instruments of state power. Carr divided political power, 'for the purpose of discussion',[73] into three categories: '(a) military power, (b) economic power [and] (c) power over opinion'.[74] Carr's categorisation of political power was significantly influenced

66 Baldwin, *Power and International Relations*. p.178.
67 Ibid. p.178.
68 See for instance: 'How Much does a Botnet Cost?' Threatpost, last modified February 28, accessed December 25, 2015, threatpost.com/how-much-does-botnet-cost-022813/77573/.
69 'Collateral Damage: 26% of DDoS Attacks Lead to Data Loss,' Kaspersky, last modified September 17, accessed December 25, 2015, kaspersky.com/about/news/business/2015/Collateral-damage-26-per-cent-of-DDoS-attacks-lead-to-data-loss.
70 Joint Chiefs of Staff, *Joint Publication 1: Doctrine for the Armed Forces of the United States* (Washington, DC: Joint Chiefs of Staff, 2013).; Ministry of Defence, *Joint Doctrine Publication 0-01: British Defence Doctrine*, 4th ed. (Shrivenham: Development, Concepts and Doctrine Centre, 2011).; Dutch Ministry of Defence, *Netherlands Defence Doctrine* (Den Haag: Ministerie van Defensie, 2013).; North Atlantic Treaty Organisation, *Allied Joint Publication 1(D): Allied Joint Doctrine* (Brussels: Nato Standardization Agency, 2010).
71 'The National Security Strategy of the United States of America,' , state.gov/documents/organisation/63562.pdf. Preface; Dutch Ministry of Defence, *Netherlands Defence Doctrine*. p.22.; Robert D. Worley, *Orchestrating the Instruments of Power: A Critical Examination of the U.S. National Security System* (Raleigh: Lulu Press, 2012). p.181.
72 Worley, *Orchestrating the Instruments of Power*. p.181.
73 Edward H. Carr, *The Twenty Years' Crisis 1919-1939: An Introduction to the Study of International Relations*, 2nd ed. (London: MacMillan & Co. Ltd., 1946). p.108.
74 Ibid.

by the *interbellum*;[75] hence parts of his categorisation may seem out-dated as he pointed out himself in the preface to the second edition. Michael Mann has forwarded a more modern approach to sources or instruments of power, namely the IEMP-model, short for ideological, economic, military, and political power sources.[76] Carr's and Mann's categorisations have been widely used and discussed in academia, but are not employed by those using the instruments or those constituting the instruments.

Although there are semantic differences between Carr, Mann and the governmental categorisations, the instruments have considerable overlaps and can be summarised as:

- the political instrument, comprising internal politics and outward facing diplomacy;
- the informational instrument, aimed at spreading, collecting, protecting and monitoring information;
- the economic instrument, comprising the use of finance and economic statecraft to influence other actors;
- the military instrument, serving as an extension of foreign policy both coercively (hard power) and attractively (soft power); and
- other (civil) capacities such as legal power, law enforcement, administrative organisations, education, healthcare, utility companies, etc.

TABLE 2: OVERVIEW OF THE INSTRUMENTS OF STATE POWER AS DESCRIBED BY CARR, MANN AND DIME(FIL) PROPONENTS, THE MEANS THESE INSTRUMENTS DESCRIBED BY VARIOUS AUTHORS AND THEIR CYBERSPACE EQUIVALENT

Instrument	Capacities	Cyber capacities
Political	• Wield other instruments (Carr) • Internal politics (Mann) • Geopolitical diplomacy (Mann) • Achieving foreign policy objectives (DIME)	• Coordination (e.g. command and control systems) • Legitimise actions via social-media • Use cyber capacities as deterrent (e.g. 0days, malware, strike-back) • Use cyber capacities to intervene abroad (e.g. Stuxnet, Orchard)
Informational	• Gain power over opinion (Carr) • Unify meaning, norms and aesthetic and ritual practices (Mann) • Controlled release of information (DIME) • Protecting own information (DIME) • Collecting information (DIMEFIL)	• Manipulate internal and external target audiences (e.g. social-media) • Shape target audiences via information campaigns (e.g. troll-army tactics) • Legitimise own and delegitimise foreign actions via (social-)media • Use cyber capacities aimed at defence (e.g. IDS, IPS, etc.) • Use cyber capacities aimed at intelligence (e.g. surveillance)
Economical	• Gain autarky or influence abroad (Carr) • Monopolise control over classes (Mann) • Support or combat other actors (DIME) • Disrupt finance of other actors (DIMEFIL)	• Protect own industries or gather competitive intelligence. • Nationalise control over Internet infrastructure/security • Support actors financially or materially (e.g. network infrastructure) • Disrupt financial traffic (e.g. SWIFT)

75 Ibid. Preface to the second edition.

76 Michael Mann, *The Sources of Social Power: A History of Power from the Beginning to A.D. 1760*, Vol. I (Cambridge: Cambridge University Press, 1986).; Michael Mann, *The Sources of Social Power: The Rise of Classes and Nation-States 1760-1914*, Vol. II (Cambridge: Cambridge University Press, 1993).; Michael Mann, *The Sources of Social Power: Global Empires and Revolution 1890-1945*, Vol. III (Cambridge: Cambridge University Press, 2012).; Michael Mann, *The Sources of Social Power: Globalisations 1945-2011*, Vol. IV (Cambridge: Cambridge University Press, 2013a).; Michael Mann, 'The Sources of My Sources,' *Contemporary Sociology: A Journal of Reviews* 42, no. 4 (2013b), 499-502.

Instrument	Capacities	Cyber capacities
Military	• Extending foreign policy (Carr) • Intensive power over limited space (Mann) • Extensive power over larger space (Mann) • Hard coercive power (DIME) • Soft attractive power (DIME)	• Intervene or stabilise abroad via military cyber capacities • Establish control over infrastructure (e.g. tailored access) • Influence actors via deterrence (e.g. a show of force in cyberspace) • Deny, disrupt, degrade, destroy infrastructure (e.g. DDoS, malware) • Information campaigns aimed at target audience (e.g. via social-media)
Other capacities	• Legal power (DIME) • Law enforcement (DIMEFIL) • Administrative organisations (DIME) • Education (DIME) • Healthcare (DIME) • Utility companies (DIME)	• Prosecuting bloggers/hackers • Arresting and detaining bloggers/hackers • Influence institutions (ICANN, ISPs, etc.) • Provide cyber awareness courses, stimulate computer sciences, etc. • Protect or deny infrastructure to opposing actors • Ibid.

Table 2 shows that most instruments have a cyber aspect to them, and that the instruments of state power extend into cyberspace. Unlike Nye, Betz and Stevens, this paper will not categorise these means under the faces of power or the taxonomy of power. The reason for not using these categorisations is that cyber capacities can be used for a wide variety of purposes, and using these categorisations could foster the thought that certain capacities can only be used for one specific purpose. That would deny the universal application of the instruments of state power and cyber capacities. All the instruments can be used to gain state objectives, military power can achieve political objectives, economic power can achieve military objectives, and (civil) capacities can attain informational objectives.

All the cyber capacities can be used for attaining any of these objectives. For instance, a DDoS aimed at a foreign military can also be used to DDoS a financial institution. Cyber capacities can exert influence via a myriad of power mechanisms, hard and soft, compulsory institutional, structural, and productive. The instruments of state power involved and their extension in cyberspace affect the effectiveness of actor A influencing actor B.

5. CONCLUSION

This paper's purpose was threefold, providing: (1) a basis for further discussion about cyber power; (2) footholds for assessing cyber power; and (3) a viewpoint on the categorisation of cyber capacities within the power debate.

Section 2 provided a brief, albeit theory-laden, overview of the 20th century power discussion. This paper, taking a relational approach to power, tried to emphasise the other dimensions of power, being: scope, domain, weight, costs, and means. Figure 1 incorporates these dimensions and mechanisms for conveying power (compulsory, institutional, structural, and productive) into a model for analysing power.

Section 3 drew the outlines of the current cyber power debate by first defining cyber power and secondly sketching the views promoted by Nye and by Betz and Stevens in their work regarding cyber power. The means they describe are very useful to anyone looking to assess cyber power, however, they have to be considered alongside the other power dimensions.

Section 4 adjoined the power analysis framework in section two with contemporary notions regarding cyber power and capacities. Any appreciation of cyber power should include the context (what is the context?); scope (what is actor A trying to achieve); domain (what actors and type of actors are involved?); weight (how likely is actor A in successfully influencing actor B?); costs (how costly is the (in)action taken by actor A and how costly is it for actor B to comply?); and means (what instruments of state power and their cyber equivalents are involved?).

To the disappointment of many policy- and decision-makers, this paper will not present a generalised overview of the powerful and the powerless actors on the world stage. The reason for not doing so is that any such statement would be flawed. A generalised overview of cyber power distribution cannot exist because it depends on all dimensions of power. As all these dimensions are contextual and some are temporal; there are too many contingencies to be captured in such an overview. This is not unique to cyber power, since the start of discussions about power it has been shown to be notoriously hard to capture in qualitative and quantitative terms.

That does not mean, however, that an appreciation of cyber power distribution is without use. An overview of the cyber power distribution, for instance by looking at the means (e.g. network infrastructure, spending, government wide cyber security budgets, acquiring malware, DDoS capacity, IT-graduates, etc.), can offer insights to policy- and decision-makers. It is, however, paramount to realise that these results have to be interpreted in their context and adjoined with the other dimensions of power. The power analysis framework forwarded in this paper can be used to that extent.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Hard Power in Cyberspace: CNA as a Political Means

Ragnhild Endresen Siedler
Analysis Division
Norwegian Defence Research Establishment
Kjeller, Norway

Abstract: This analysis is a contribution to the scholarly debate on how cyber power influences international relations. It answers the following question: In what ways can an actor apply CNA to dominate an opponent, and what may restrict him in this effort? It uses Schelling's (2008) argument for dividing power into coercion and brute force, and thus the paper distinguishes between actions that inflict harm and those that impose limitations. Through this approach, it describes the difference between CNA as a means of pure destruction and CNA as a means of forcible accomplishment in order to elucidate different ways of using CNA. This analytical approach aims at generating insight into the nature of CNA threats, which in turn, facilitates development of appropriate responses. The paper argues that defensive cyber strategies and doctrines primarily should focus on CNA as a means of forcible accomplishment. However, it also discusses CNA in the form of coercive threats. It explores this type of power by assessing how the technological and organizational preconditions of CNA apply to severity of consequences and credibility. In particular, two challenges make such threats less credible: unpredictability of consequences, and the ability to make the defender expect that there are more attacks to come. However, one coercive potential of CNA may be in the form of reprisals.

Keywords: *cyber power, hard power, computer network attack*

1. INTRODUCTION

This paper analyzes different ways of using Computer Network Attack (CNA) as a means of power in the context of interstate conflict. Several sources argue for the necessity to develop appropriate responses to hostile actions in cyberspace (U.S. Department of Defense, 2012; NATO, 2014a; Hopkins, 2013). Appropriate responses, however, require insight into the nature of the hostilities in question. In media reports, cyber threats are often described in sensational terms (for example, in Goldman, 2012). However, conveying nuances and an in-depth understanding of the potential power of cyber threats is crucial for identifying appropriate responses.

Here, the term cyber hostility is used as a general term for activities in cyberspace that are in pursuit of an actor's interests and detrimental to his/her opponent. Nye (2004, p.5; 2011, p.11) distinguishes between hard and soft power. Whereas soft power "(...) rests on the ability to shape the preferences of others" (ibid, 2004, p.5), hard power influences peoples' behavior by sanctions. Negative sanctions ("sticks") punish undesired behavior and positive sanctions ("carrots") reward desired behavior. This paper centers on CNA only, and how CNA may be applied to shape circumstances to an attacker's[1] advantage[2] and thus falls into the category of hard power behavior. Hence, the research question is:

> *In what ways can an actor apply CNA to dominate an opponent, and what may restrict him in this effort?*

To answer this question, it may be useful draw on Schelling's (2008) conceptualization of power as either brute force or coercion. This paper distinguishes coercion from brute force according to Schelling's (2008) theoretical argument. Often, CNA threats will occur concurrently, or even in conjunction with, other hostile actions. However, before the interplay between CNA and other actions can be fully explored, it is necessary to gain in-depth understanding of what CNA actually is as an independent means. For the sake of scope and focus, this paper focuses on the latter aspect only. First, it defines how cyberspace relates to the phenomenon of power in the political science sense of the term. Second, it describes how cyberspace's technological attributes constitute preconditions for the political utility of CNA. Finally, it discusses how CNA applies to coercion and brute force respectively.

This paper argues that CNAs shape circumstances mainly by imposing limitations rather than by inflicting harm. CNAs may succeed in being obstacles for military and governmental crisis management, but as a means of coercion, there are critical challenges with respect both to its credibility and to the severity of its consequences.

2. POWER

There is a distinction between the meaning of "cyber power" in political science and how the same expression is understood in military studies. The essence of political science is the analysis of "who gets what when and how" (Lasswell, 1936). In the military context, by contrast, "cyber power" is equivalent to military power projection in the physical domain.[3] This paper discusses cyber power in the political science sense of the term.

From a realist perspective, power is a resource that can be distributed between actors, and thus power is a relative phenomenon in which the enhancement of the power of one actor means a consequent loss to his opponent (Waltz, 2001, p.210; Jackson and Sørensen, 2007, pp.45–46). In the context of conflict, the ability to dominate is a key aspect. Thus, cyber power can enable an actor to dominate opponents. Moreover, conflicts arise as a result of conflicting interest. In

1 I focus on two types of actors: the attacker, who is that actor applying CNA, and the defender, the attacker's opponent and the victim of attack.

2 Computer Network Exploitation (CNE) is also a form of cyber hostility, but, in contrast to CNA, CNE is a type of intelligence collection and not a political means to dominate other actors directly. Hence, CNE is not included in this analysis.

3 For example, the U.S. Department of Defense's (2015) definition of maritime power (projection) as "Power projection in and from the maritime environment, (…)".

such situations, the ability to pursue interests despite actions from opponents is a central aspect. Thus, cyber power implies an ability to achieve one's own interest.

In order to achieve interests, however, power requires actions to transform its potential into results, namely the exercise of power. Hoffman and Graham (2006, p.4) argue that a state of freedom is a precondition for exercising power, and that freedom, in turn, implies having an influence on one's surroundings. Deriving from this logic, cyber power is a phenomenon that empowers an actor to shape circumstances into his/her advantage through actions in or via cyberspace, and thus to pursue self-interests through such actions.[4]

2.1 Schelling: What force can do

Schelling is frequently used by political and military scholars to understand how military means can rationally be applied in pursuit of security objectives. He describes how arms can enable an actor to exercise power. He separates force – how arms influence actors in conflict – into two types: coercion, and brute force. The former is "the power to hurt" and the latter "the power to seize or hold forcibly." Coercion is a latent aspect, which implies that it does not require the use of any means in order to be effective. On the contrary, coercion is most effective when actors comply without any actions being physically carried out (Schelling, 2008, p.10).

Coercion is the main focus in Schelling's (2008) analysis.[5] He argues that this is a form of bargaining power that actors (he mainly focuses on state actors) make use of in order to affect other actors' cost-benefit analysis. This way, coercive power changes and shapes decision-making, and thus is strategic in nature. The ability to influence decision-making, however, is closely related to the ability to inflict harm, or, in the words of Schelling (2008), "the power to hurt":

> "(...) the sheer unacquisitive, unproductive power to destroy things that somebody treasures, to inflict pain and grief–is a kind of bargaining power (...)." (Ibid, p. xiii).

Any threat of harm must be credible. According to Schelling (ibid, p.35), credibility is about communicating that an actor actually intends to carry out hostile actions, as opposed to leaving the impression that it is a bluff. Schelling's focus on strategic decision-making implicitly assumes that such a capability is in place. With respect to cyberspace, however, the technological requirements (as described in the next section) are essential for an attacker's ability to conduct CNA operations, and in the absence of displayed weapons, this paper discusses credibility in the light of what is technologically possible and how that influences the extent to which a CNA

4 There are several definitions of cyber power. Czosseck (2013), for instance, defines this phenomenon as "(…) the ability to act and influence through, and by means of, cyberspace." (Ibid, p.1). Nye (2011, p.123) emphasizes the resource aspect of power and the ability cyberspace provides to gain preferred outcomes. Despite differences in wording, the definition of this paper does not conflict particularly with other relevant definitions of the same term. In this analysis, however, the aspect of "shaping circumstances" is a central aspect. Hence, I use the definition above.

5 Schelling (2008) elaborates on how coercion can be applied strategically in different ways. The distinction between brute force and coercion is theoretical, and the starting point for further elaboration on coercion. The theory developed during the cold war, when strategic air power and nuclear weapons were predominant capabilities. A question, therefore, is whether this provides a useful theoretical basis for analysis of international relations today, and particularly whether it is relevant as an analytical perspective for informing the understanding of cyber power. Despite the cold war context, I argue that Schelling's (2008) theoretical perspective still provides explanations of general value for how arms influence conflicts, and thus may help to elucidate the nature of new types of weapons, including CNA.

threat is convincing.[6] Credibility depends on the defender's expectations (Ibid, p.56). One way in which an actor can enhance his/her credibility is through the application of capabilities. In addition to inflicting "pain and grief" in a specific incident, he can demonstrate that there will be more of such consequences to come, or, as Schelling (2008) puts it when commenting on the nuclear bombing of Japan in 1945: "They hurt, and promised more hurt, and that was their purpose." (Ibid, p.17).

The other type of power, "brute force," is defined by Schelling (2008) as the application of strength in pursuit of self-interest *without* appealing to any other actor's decision-making. It has a kind of "just do it" character. It is the application of capabilities to simply enforce or carry out actions in pursuit of a given objective. Schelling (2008) characterizes this form of power with three key words: strength (in relative terms), skill, and ingenuity. Typical objectives are, for instance, to seize, to penetrate, to occupy, to disable, and to deny access (ibid, p.1). The following quote summarizes the essence of the distinction between brute force and coercion:

> "There is a difference between taking what you want and making someone give it to you (…) between losing what someone can forcibly take and giving up to avoid risk or damage." (Ibid, p.2).

To summarize, brute force is the power to hold or seize forcibly, and CNA in this form of power would be a means of forcible accomplishment, which I in turn interpret as a way of imposing limitations. Coercion, by contrast, is the power to hurt, and CNA in this context would be a means of pure damage, which in turn implies the infliction of harm (ibid, p.8). Table 1 summarizes the distinction between coercion and brute force.

TABLE 1: HOW COERCION DISTINGUISHES FROM BRUTE FORCE

Coercion	Brute force
Inflict harm	Impose limitations
Pure destruction or damage	Forcible accomplishment
The power to hurt	The power to seize or hold forcibly
Coercion is a latent aspect	
A form of bargaining power	
Coercive power changes and shapes decision-making	Application of strength without appealing to any other actor's decision-making

Following Schelling (2008), this paper distinguishes between actions that impose limitations on an actor's capability and those that inflict harm.[7] Deriving from Schelling's (2008) theoretical argument, hard cyber power can be exercised either to influence opponents' capabilities or their

6 Displaying a CNA capability would indirectly point out how a defender could neutralize the threat, as described in more detail in section 4. The contradiction between secrecy of capabilities and the rationality in "show of force," is also discussed in section 4.

7 Correctly, brute force can imply more than mere restrictions or obstacles, as, for instance, in cases of occupation. However, in order to emphasize the contrast with harm, this paper uses the word "limitations," which involves restrictions on a defender's capabilities to a greater or lesser extent, encompassing the whole spectrum of outcomes from disablement and denial of access, to occupation.

decision-making via actions in cyberspace. The next sections elaborate in more detail what such actions in cyberspace can be, and link these actions to their ability to pursue interests.

3. WHAT IS CNA EXACTLY AND WHAT DOES IT MEAN FOR POLITICAL UTILITY?

This paper uses NATO's definition of CNA:

> "Action taken to disrupt, deny, degrade or destroy information resident in a computer and/or computer network, or the computer network itself." (NATO, 2014b).[8]

Maybaum (2013) presents an explanation of what steps a CNA operation consists of, and what consequences it may cause.[9] Several sources distinguish between sophisticated and unsophisticated CNA, based on level of capability requirements (Nye, 2010, p.11; Libicki, 2009, p.154). By sophisticated CNA, this paper means operations that are resource- and intelligence-intensive, as described later in relation to the Cyber Kill Chain; that to a large extent exploit logical vulnerabilities such as zero-day vulnerabilities;[10] and that generally require more effort to circumvent security measures.[11] Less advanced CNAs, by contrast, make use of techniques that are easier to access, such as Denial of Service (DoS) attacks, which could be referred to as digital jamming. Such techniques are often based on off-the-shelf attack tools available on the free market. Though the method of producing a tool might be sophisticated, the attacker can easily apply them without any significant level of expertise.[12] Lindsay (2013) supports a distinction based on sophistication: "The difference in scale between strategic cyber warfare and cheaply available cyber irritants will become more apparent in the matter of offense dominance." (Ibid, p.389).

The level of sophistication an attacker has determines what targets he is able to engage (Libicki, 2009, p.154). Due to the low level of capacity needed for unsophisticated CNA, many actors can acquire a credible but unsophisticated CNA capability. The question is, however, whether that capability is sufficient to influence other actors' capabilities or decision-making.

A sophisticated CNA operation can be described as an array of seven steps: (1) reconnaissance, (2) weaponization, (3) delivery, (4) exploitation, (5) installation, (6) command and control and, finally, (7) actions on objectives (Hutchins et al., 2011). This array is referred to as the Cyber

8 Several sources use other expressions, such as "cyber attack" or "cyber operation." These terms may be confusing, as some include CNE in these terms. For the sake of clarity, this paper consistently uses the term CNA with the definition given above.

9 Maybaum (2013) bases his argument on Hutchins et al. (2011), which is the original source for the Cyber Kill Chain. Because Maybaum's (2013) chapter specifically presents the technologically heavy aspects of the Cyber Kill Chain for a non-technological audience, it is more applicable for this paper, and I mainly use that source here.

10 Stuxnet is frequently mentioned as a highly sophisticated CNA operation. It exploited four zero-day vulnerabilities (Lindsay, 2013). A zero-day vulnerability is defined as "(…) a hitherto undisclosed, exploitable vulnerability (…)." (Czosseck, 2013, p.12).

11 The description of the technological attributes mentioned in this paragraph derives from a meeting with the Norwegian Defence Research Establishment's cyber scientists Ronny Windvik, Torgeir Broen and political scientist Torbjørn Kveberg, Kjeller, 27 November 2015.

12 An alternative to exploiting logical vulnerabilities is social engineering, tricking users to obtain passwords and usernames to access a system or website (Norwegian Defence Research Establishment's scientists, as in footnote 11).

Kill Chain (CKC). Maybaum (2013, p.130) describes the CKC as "a very complex endeavour" in which the intruder is dependent on "deep system knowledge at expert level." This is why sophisticated CNA operations are resource intensive, time consuming, and critically dependent on intelligence. Thus, a sophisticated CNA is difficult to conduct within a short timeframe.

In addition, it is reasonable to expect that the more valuable an asset is to the defender, the more security measures will be in place to protect it. This adds even more challenges for the CNA–developer, and the level of sophistication will have to increase accordingly. Sustaining access to a target over time and awaiting the optimal time of engagement also requires additional effort (Maybaum, 2013, pp.122-126). This restricts how many high-value targets an attacker can simultaneously attack.[13]

Moreover, the malicious code has to be uniquely designed for the target in question (Maybaum, 2013, p.112). This implies that, in contrast to conventional weapons and weapons of mass destruction, "cyber weapons" (or "cyber tools,") used in sophisticated CNAs cannot be stockpiled in huge numbers and used against whatever target emerges at short notice. Hutchins et al. (2011) argue that the defender can conduct several actions to outmaneuver an attacker, and that this is a key advantage in favor of the defender. In other words, the defender too can shape the circumstances to his advantage.

4. COERCIVE POWER IN CYBERSPACE

This section discusses to what extent CNA can influence others' decisions by serving as a "means of pure damage." This implies a discussion of to what extent it can inflict harm. It elaborates on the severity of consequences and then addresses aspects of credibility.

4.1 Severity of consequences

In the public discussion, civilian targets in cyberspace are often conflated with military and governmental crisis management targets.[14] However, in order to make the distinction between harm and limitations more tangible, I distinguish civilian targets from military and governmental crisis management targets. Schelling (2008, pp.8-9) emphasizes the need to specify the level of decision-making being subject to an attacker's influence; that is, whether it is an individual or the strategic decision-maker. As this paper centers on interstate conflicts, it focuses on the level of strategic decision-making. Whereas military and governmental crisis management is by nature instrumental and relates to capabilities, civilians do not represent a capability. Instead, civilians represent that part of an actor where harm materializes as suffering. The rationality for targeting non-capabilities, logically, is not to influence the relative strength between two competitors, but the cost-benefit consideration of the strategic decision-maker and his/her

13 Norwegian Defence Research Establishment's scientists, as in footnote 11.

14 In cyberspace, military and governmental crisis management targets can be information channels between authorities and the population, or communication systems internally in the government to exchange information and coordinate actions. Thus, this can be information infrastructure underpinning the military's or the government's ability to handle the situation effectively.

motivation to avoid risk or damage in the bargaining over competing interests (Schelling, 2008, p.2).[15]

Several scholars (Libicki, 2009; Geers, 2010; Rid, 2013) argue that the impact of CNAs on humans in comparison to other hostile actions such as kinetic attacks and, ultimately, nuclear weapons, are of a low scale. Geers (2010) summarizes this argument: "Cyber attacks *per se* do not cause explosions, deadly heat, radiation, an electro-magnetic pulse (EMP), or human casualties." (Ibid, p.301). This comparison with the consequences of other means of attack is important because it places the effects of CNA in perspective with those of the alternative options an attacker may have at his/her disposal.

An example in which the consequences of CNAs are described in severe terms (for instance, Ottis, 2008), is the Estonia 2007 case. Here, less sophisticated CNA disrupted both private and governmental targets (Tikk et al., 2010).[16] At the same time, there was ongoing disagreement over the movement of a Soviet World War II memorial. Hence, these actions can be interpreted as attempts to influence a political decision.[17] The logic in Schelling's rational choice argument of coercive power is that the costs of giving in must be lower than the costs of suffering a given hostility. Thus, severity is a relative phenomenon. Whether something is considered as severe or not must be regarded in the light of the disputed issue at stake. The cyber hostilities disrupted private businesses and led to economic loss. In addition, they hampered governmental administration services – according to Tikk et al. (2010), a matter of societal security – and Estonia's international communication channels (ibid). Although these actions were described as "the worst-ever DDoS" at that time (Rid, 2013, p.6), Rid (2013) concludes that, although noticeable, the effects remained minor. Tikk et al. (2010, p.25), however, classify the effects as "beyond mere inconvenience," and Ottis (2008, p.2) goes even further, categorizing the situation as a "threat to national security." These assessments of the same consequences illustrate that perception of impact varies to a certain extent.

Despite some variation in perceptions, the question in relation to Schelling's "means of inflicting pure damage," is whether the consequences qualify as sufficient "pain and grief" to influence decision-making. Although the actions resulted in a high number of attacks across Estonian society, encompassing both civilian and governmental targets, and although the effects were noticeable, there are, to the best of my knowledge, no reports of human suffering per se.[18] Kinetic attacks, by contrast, could have led to such outcomes. Therefore, I argue that the consequences in the Estonia 2007 case do not qualify as the severity level of "pain and

15 The principle of distinction in the Geneva Convention, Additional Protocol 1, (International Committee of the Red Cross, 1977) prohibits actors from making civilians the objective of armed attack. This is an indication in itself of the level of severity of such actions, and it also represents a legal restriction on the use of CNA. Notably, the distinction between capabilities and civilians in this analysis is based on Schelling's (2008) distinction between brute force and coercion, and not on legal regulations.

16 Methods used were DoS attacks, Distributed Denial of Service (DDoS) attacks, defacing, attacks on the Domain Name System server, and e-mail and comment spams (Tikk et al., 2010, p.33).

17 There is some uncertainty as to whether the attacker was a state or a number of non-state actors (Rid, 2013, p.7). Due to the scope of interstate conflicts, however, this analysis assumes that the attacker was a state actor.

18 One particular target is worth noting, namely the temporary disruption of the national emergency number, 112 (Tikk et al., 2010, p.21). If citizens in need of public emergency services had not received critical assistance, the consequences could potentially have resulted in human suffering. However, Tikk et al. (2010) emphasize that this number was only briefly blocked, and severe consequences are not reported in this source.

grief" compared to the defender's stake in the conflict, and that the hostilities therefore did not provide the attacker with enough "power to hurt." Consequently, I support Rid's (2013, p.6) statement that, in the end, the attacks did not lead to any significant change in the status quo. Hence, this serves as an example of unsophisticated CNA that did not succeed in producing enough harm to coerce the defender into concessions. If the unsophisticated CNAs on Estonia had any coercive effect at all, it may have been in the form of reprisals. Instead of changing a current decision, reprisal aims at influencing future decisions by imposing a form of punishment (Schelling, 2008, p.79). Although the status quo did not change in the dispute over the World War II memorial, the attacks may have succeed in making resistance costly enough to make both this and future opponents think twice before they again act in a way that is contradictory to the attacker's interests.

In sum, the impact of unsophisticated CNA on people, by comparison to kinetic attack, is relatively low. As an independent tool, unsophisticated CNA faces challenges in making the defender give up, despite civilians being among the victims. However, the coercive impact on future decisions in the form of reprisals may be more relevant.

4.2 Credibility

The threat of sophisticated CNA targeting Industrial Control Systems (ICS) is frequently mentioned in public debate (for instance by Rid, 2013, p.66). The reason is that this kind of control mechanism is almost ubiquitous: in power plants, water plants, electrical and transportation grids, traffic lights, and hospitals, to name but a few examples (ibid, p.67). This dependency represents a source of vulnerability that an attacker could exploit to inflict physical disruption or damage.

Nye (2011) suggests a hypothetical example of a CNA on a civilian target:

> "If a hacker or a government shut down the provision of electricity in a northern city like Chicago or Moscow in the middle of February, the devastation could be more costly than if bombs had been dropped." (Ibid, p.127).

The quote implies that the consequences of such an attack could inflict harm by causing civilian suffering. Given that such a scenario would be technologically and organizationally possible, and assuming that it would be sufficiently severe, would the threat be credible? I argue that there are two main challenges: (1) unpredictability of consequences; and (2) the ability to make the defender expect that there are more attacks to come.

Libicki (2009, p.79ff) argues that, because such attacks depend on exploiting the defender's vulnerabilities, their consequences are more difficult to predict than those of physical attacks.[19] Unpredictability of consequences makes it difficult to convince the defender that he would be better off giving in. What risk or damage is he actually going to avoid?[20] The fact that consequences are difficult to predict may also weaken the defender's belief in the attacker's willingness to use a means in which the outcomes are uncertain. It would also make it less

19 Additionally, the complexity in such systems makes it difficult to predict the consequences of actions (Norwegian Defence Research Establishment's cyber experts, see footnote 11).

20 Arguably, uncertainty is a factor that can increase the perception of a specific risk. However, the type of uncertainty discussed here is rather of a kind that contributes to enhancing doubt instead of fear of consequences.

convincing that the consequences would be sufficiently severe compared with the stakes in question. Contrasting Schelling's (2008) claim that the use of capabilities enhances credibility, Libicki (2009, pp.79-80) argues that, by demonstrating an example the attacker simultaneously reveals the vulnerability on which he bases his CNA operation, and thus indirectly points out how the defender can shape the circumstances to his/her advantage by reinforcing cyber security. When in place, these measures make the attack in question even less convincing as a "promise of more," and the defender is less likely to give in to the attacker's demands (ibid, pp.79-80). Therefore, attempting to achieve coercive power through sophisticated CNA by such a "show of force" is highly challenging.

The uniquely designed code needed for this kind of action also makes it difficult to achieve a convincing "promise of more" expectation. In contrast to air strikes, an actor conducting CNA to take out the centrifuges in the nuclear facility in Natanz does not automatically have the same capability to take out centrifuges in another facility, at least not within the short time required to create a coercive power potential in an ongoing conflict. Arguably, succeeding in such an advanced operation demonstrates a capacity in both organizational and technological terms, which in turn substantiates the possibility of similar actions. However, although the attacker can take some shortcuts, a similar attack would require the development of a new and bespoke CKC designed for the new target. Additionally, the time, resources, and willingness to accept risk of exposure needed for accomplishing all seven steps of the CKC limits how many CKCs an attacker can simultaneously conduct during the course of a conflict. This, therefore, also makes it challenging to produce a "promise of more" expectation in the defender.

Another aspect is the attribution problem, which is often presented as the attacker's key advantage in cyberspace (Rid, 2013, p.139ff).[21] With respect to coercion, however, hiding one's identity is not necessarily an advantage. Libicki (2009) makes the same point: "To coerce, an attacker must signal that a specific set of cyberattacks was meant to coerce." (Ibid, p.80). Instead of denying, he has to *prove* responsibility. Credibility is a factor expressing to what extent the threat of a hostile action is convincing, that is, whether the attacker succeeds in persuading or dissuading an opponent. Logically, in persuading an opponent to give in, it must be convincing that, if he does, the attacker will not inflict the harm the defender gives in to avoid. This is what Schelling (2008, p.74) refers to as "corresponding assurances." If the defender is uncertain of the attacker's identity, how can he assess how the attacker would act in response to his decisions? Therefore, confusing a defender is somewhat counterproductive to the objective of convincing him. Instead, confusion is a form of imposed limitation on crisis management, and can thus be seen as brute force rather than coercion.

These arguments illustrate that a "promise of more" effect is challenging to achieve through sophisticated CNA for two reasons: the tool has to be uniquely designed; and use of such attacks would point out where to close security gaps. In contrast to the use of brute force, the attacker would find it difficult to make use of the attribution problem to his advantage when coercing. If the identity is obscure, the attacker would find it difficult to convince the defender of "corresponding assurances." Instead, the defender may perceive all these obstacles to successful sophisticated CNAs as indications that such threats are bluffs rather than realities.

21 The attribution problem refers to difficulties of tracing attackers in cyberspace.

5. CYBERSPACE AND BRUTE FORCE

How can CNA serve as a "means of forcible accomplishment?" Schelling (2008) summarizes the essence of such a means in three key words: strength, skill and ingenuity. The CKC in itself is interesting. As Maybaum (2013) points out, it requires an optimal match of many details to succeed in all seven steps of the CKC. This, in turn, implies a high threshold for success, particularly for high-value targets. Thus, an attacker will be greatly in need of strength in terms of technological sophistication, skill and ingenuity, in its reconnaissance, social engineering, and ways of accessing networks.

Lindsay's (2013) analysis of the Stuxnet case provides a practical example. He describes how demanding the Stuxnet operation was throughout the different steps of the CKC. His analysis demonstrates that the operation required detailed information on a number of different factors and a highly qualified team with diverse expertise. Hence, sophisticated CNA on high-value targets is an operationalization of the three key words of brute force in their own right, or in the words of Lindsay (2013, p.379): "Stuxnet's technical particulars reveal a devious genius on the part of its developers (…)."

Stuxnet illustrates how a CNA reportedly resulted in physical consequences, in that the speed of the centrifuges in Natanz was manipulated (Lindsay, 2013).[22] The operation is discussed in relation to the international conflict over the Iranian nuclear program and its possible military dimensions. In a nuclear program, enrichment facilities are critical assets. Stuxnet was a covert action that exploited the attribution problem to achieve its objective (ibid). By this, Stuxnet represents a contrast to the exercise of coercive power, which, as argued in relation to credibility, requires proven responsibility instead of deniability. The fact that the attacker did not explicitly appeal to the defender's decision-making (in relation to this particular action), but instead forcibly imposed limitations, underscores that Stuxnet is an example of brute force. Moreover, it is an empirical example of disablement – one of the actions Schelling proposes as types of brute force.

Whereas Stuxnet serves as an example of sophisticated CNA at the strategic level (Lindsay, 2013), Libicki (2009, p.139ff) proposes several ways for how CNA can be used at the operational level of warfare. In particular, he emphasizes the advantage CNA has in taking a defender by surprise. Surprise is a condition imposed on adversaries to gain initiative and make operational advances, or, as Joint Publication 3-0 (U.S. Department of Defense, 2011, p.A-3) puts it: "Surprise can help the commander shift the balance of combat power and thus achieve success well out of proportion to the effort expanded." Therefore, actions causing surprise may have a favorable return compared with their costs. Additionally, disruptions which temporarily incapacitate the defender may provide an attacker with a valuable window of opportunity (Libicki, 2009, p.146). Assuming that such actions are technologically possible, examples of operational use could include blinded sensors, misinterpretations of signals, and weapon system malfunctions (ibid, p.139ff). These are other examples of how CNA can serve as acts of disablement.

22 To what extent the attacker actually succeeded in imposing limitations in the Stuxnet case, however, has been questioned (Lindsay, 2013), but elaborating on this aspect further is beyond the scope of this paper.

Because power is a relative phenomenon, enhancing one actor's "power to seize/hold forcibly," decreases his/her opponent's power to resist or conquer. Therefore, imposing limitations on the opponent's capabilities as illustrated here represents advantages that the attacker can exploit to accomplish his mission.

In 2008, an unsophisticated CNA hit Georgia concurrently with the armed interstate conflict between the country and Russia over the geographical area of South Ossetia. Specifically, this consisted of defacement of key websites, DDoS attacks on government and civilian targets, and spamming of politicians' e-mail accounts (Tikk et al., 2010, p.67ff). The methods used were similar to those seen in Estonia in 2007 (ibid, p.71). However, one noticeable feature of the Georgia case was the targeting of online government communication channels. Tikk et al. (2010) comment on the implications in this way:

> "The Georgian government, reliant on its official websites as information distribution channels, had to look for ways of avoiding information blockade." (Ibid, p.70).

Instead of interpreting this action as an appeal to Georgia's political leadership to change its position in the conflict, the hostilities can be seen as attempts to disrupt its ability to manage the situation, thereby reducing its ability to shape circumstances in favor of its interests. In conflicts, efficient crisis management is a competition between two rivals and their respective decision cycles (Boyd, 1976).[23] Using CNA to disrupt decision cycles, as in the Georgian case by attempting to disrupt communication, and thereby hampering execution of decisions may serve as a useful way to impose limitations on an opponent.[24]

The examples above illustrate how an attacker can use a technological edge to shape circumstances into his advantage without directly harming the defender to concessions. Instead, he uses his strength, skill and ingenuity in cyberspace to "just do it" without appealing to decision-making. This supports the argument that CNA can be used as a "means of forcible accomplishment."

6. CONCLUSION

This paper demonstrates that it is useful to distinguish between CNAs that aim to impose limitations, and those attacks that aim to inflict harm. Sorting threats into these two categories would help analysts and decision-makers to better understand a situation involving a CNA threat. For this reason, the paper also distinguishes between civilian targets and military and governmental crisis management targets.

This paper finds that CNA has the potential to dominate an opponent in the form of "power to seize/hold forcibly," but as "power to hurt" there are challenges with both communicating credibility and causing sufficiently severe consequences. Two aspects represent restrictions in

23 Boyd's (1976) decision-theory resulted in a model describing how organizations conduct decision-making in a circle of four steps: Observe – Orient – Decide – Act, also referred to as the OODA loop.

24 According to Rid (2013, p.8) this attack had limited effect because the Georgian government moved the targeted channel over to a foreign service provider. Despite limited effects in this particular case, it provides a practical example for how CNA potentially can be applied as a "means of forcible accomplishment."

making CNA threats credible. First, uncertainty of consequences makes it difficult to convince the defender that he/she is better off avoiding the risk. Second, it is difficult to convince the defender that there is a "promise of more" of such actions. A more likely coercive potential of CNA is in the form of reprisals, making this and future opponents think twice next time they consider acting in conflict with the attacker's interests.

As a consequence, defensive cyber strategies and doctrines should primarily focus on CNA in the form of a "means of forcible accomplishment," however, keeping in mind that CNA may have a coercive potential in the form of reprisals. Additionally, the results of this analysis may help analysts to nuance the threat picture, and in turn enhance situational awareness among decision-makers.

In relation to the overall phenomenon of hard cyber power, this analysis illustrates how means of power in cyberspace can be applied to influence opponents in the context of interstate conflicts. Given that actions are technologically and organizationally possible, the analysis indicates that hard cyber power primarily has the potential to influence opponents in the form of brute force.

This analysis also illustrates that Schelling's theoretical perspective, which was developed in the context of the cold war and the threat of nuclear weapons, is of general value in facilitating enhanced insight into new types of arms, and even into CNA. Finally, this paper has analyzed CNA as an independent means, omitting the interplay with other actions and how this interplay may change the impact of CNAs. Future work may therefore find it useful to explore this aspect in more detail. Analysing non-state actors and intra-state conflicts in light of Schelling's (2008) theory would also be interesting issues for future research on cyber power.

ACKNOWLEDGMENTS

I would like to express my gratitude to the Norwegian Defence Research Establishment's (FFI) scientists Ronny Windvik, Torgeir Broen and Torbjørn Kveberg for useful and clarifying inputs on technological issues and cyber operations. Additionally, I would like to thank FFI scientist Dr. Tore Nyhamar for advice on political analysis, and for encouraging me to develop this idea into a conference paper. However, the responsibility for any errors or omissions is mine alone.

REFERENCES

Boyd, John R. 1976. "Destruction and Creation." Version 3, September 1976. Available: http://goalsys.com/books/documents/DESTRUCTION_AND_CREATION.pdf. Accessed: 9 December 2015.

Czosseck, Christian. 2013. "State Actors and their Proxies in Cyberspace." In Ziolkowski, Katharina (ed.), *Peacetime Regime for State Activities in Cyberspace. International Law, International Relations and Diplomacy*. Tallinn: NATO CCD COE Publication, pp.1–29.

Geers, Kenneth. 2010. "The Challenge of Cyber Attack Deterrence." *Computer Law & Security Review* 26(3), pp.298–303.

Goldman, David. 2012. “Major Banks Hit with Biggest Cyberattacks in History.” *CNN*, 28 September 2012. Available: http://money.cnn.com/2012/09/27/technology/bank-cyberattacks/. Accessed: 14 December 2015.

Hoffmann, John and Graham, Paul. 2006. *Introduction to Political Theory*. Harlow, Essex: Pearson Education Limited, 1st edition.

Hopkins, Nick. 2013. “British Military at Risk of ‘Fatal’ Cyber-Attacks, MPs Warn.” *The Guardian*, 9 January 2013. Available: http://www.theguardian.com/politics/2013/jan/09/armed-forces-cyber-strategy-criticised. Accessed: 14 December 2015.

Hutchins, Eric M., Cloppert, Michael J. and Amin, Rohan M. 2011. “Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains.” Paper presented at the 6th International Conference on Information warfare and Security, George Washington University, Washington, DC, 17–18 March 2011. Available:http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf. Accessed: 28 October 2015.

International Committee of the Red Cross. 1977. “Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I).” Available: https://www.icrc.org/applic/ihl/ihl.nsf/vwTreaties1949.xsp. Accessed: 8 December 2015.

Jackson, Robert and Sørensen, Georg. 2007. *Introduction to International Relations – Theories and approaches.* New York: Oxford University Press, 3rd edition.

Lasswell, Harold D. 1936. *Politics – Who Gets What, When, How*. New York: McGraw-Hill.

Libicki, Martin C. 2009. *Cyberdeterrence and Cyberwar*. Santa Monica, CA: RAND Corporation.

Lindsay, Jon R. 2013. “Stuxnet and the Limits of Cyber Warfare.” *Security Studies*, 22(3), pp.365–404.

Maybaum, Markus. 2013. “Technical Methods, Techniques, Tools and Effects of Cyber Operations.” In Ziolkowski, Katharina (ed.), *Peacetime Regime for State Activities in Cyberspace. International Law, International Relations and Diplomacy*, Tallinn: NATO CCD COE Publication, pp.103–131.

NATO. 2014a. “Wales Summit Declaration.” Available: http://www.nato.int/cps/en/natohq/official_texts_112964.htm?mode=pressrelease. Accessed: 9 March 2016.

NATO. 2014b. “NATO Glossary of Terms and Definitions.” (AAP-06 (2014)). NATO Standardization Agency. Available: http://nso.nato.int/nso/zPublic/ap/aap6/AAP-6.pdf. Accessed: 28 October 2015.

Nye, Joseph S., Jr. 2004. *Soft Power – The Means to Success in World Politics*. New York: PublicAffairs 2004, 1st edition.

Nye, Joseph S., Jr. 2010. *Cyber Power*. Boston: Harvard Kennedy School, Belfer Center for Science and International Affairs.

Nye, Joseph S., Jr. 2011. *The Future of Power*. New York: PublicAffairs 2011, 1st edition.

Ottis, Rain. 2008. “Analysis of the 2007 Cyber Attacks Against Estonia from the Information Warfare Perspective.” Tallinn: CCD COE. Available: https://www.etis.ee/Portal/Publications. Accessed: 29 October 2015.

Rid, Thomas. 2013. *Cyber War Will Not Take Place*. London: Hurst & Company, 2013.

Schelling, Thomas C. 2008 [1966]. *Arms and Influence*. New Haven, CT: Yale University Press. 2008 edition.

Tikk, Eneken, Kaska, Kadri and Vihul, Liis. 2010. *International Cyber Incidents – Legal considerations*. Tallinn: NATO CCD COE.

U.S. Department of Defense. 2011. “Joint Operations.“ Joint Publication 3-0 (JP 3-0).

U.S. Department of Defense. 2012. “Joint Operational Access Concept.” Version 1.0, 17 January 2012.

U.S. Department of Defense. 2015. “Dictionary of Military Terms.” Available: http://www.dtic.mil/doctrine/dod_dictionary/. Accessed: 17 December 2015.

Waltz, Kenneth N. 2001 [1959]. *Man, the State and War – A Theoretical Analysis*. New York: Colombia University Press. 2nd edition.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Winning and Losing in Cyberspace

Jason Healey
Saltzman Institute of War and Peace Studies
Columbia University SIPA
New York, NY, USA
jh3639@columbmbia.edu

Abstract: This paper examines cyber conflict using a lens of 'winning' or 'losing', or alternatively the role of 'victory' or 'triumph' compared to that of 'defeat', to draw broader conclusions about the dynamics of cyber power. To do so, the paper analyses writing on the general topic of winning over the years, then dives into the two most critical key case studies: the 2007 attacks on Estonia, and the 2008-2015 conflict between the United States and Iran. It addresses the most relevant factors for these cases, including a summary of the participants in the conflict and which side 'won' or 'lost' and why. After these case studies, the paper will address larger questions of winning and losing and the implications for our understanding of cyber power.

One of the factors that most distinguishes this research from previous work on cyber power is that winning is defined not only by actions on the network, but in terms of longer-term national security outcomes. For example, Estonia certainly lost tactically in 2007, as it was offline because of the Russian-encouraged denial-of-service attack. Regretfully, most analyses of the conflict do not explore any further, which is unfortunate, as while the Estonians lost the battle, they won the war. The Estonians refused to be coerced and are now renowned for their cyber security excellence, while NATO was warned of the dangers of cyber conflict, even building a new NATO cyber centre of excellence in Tallinn. Russia was thereafter known as a cyber bully. When expressed in terms of longer-term national security outcomes, it is clear they won both operationally and strategically.

Because this larger, non-technical view is often ignored in analyses of cyber conflict, this paper makes the case that the United States and nations that follow its model misunderstand the dynamics of cyber power and cyber conflict. Too much emphasis is placed on the success or failure of offensive and defensive capabilities, rather than on better or worse long-term national security outcomes.

The paper concludes with a short view of what winning might mean in more strategic national

security terms, and recommendations for the mechanics of national security policy-making.

Keywords: *cyber conflict, cyber power, case study, Estonia, Iran, winning, defeat, losing, victory*

1. INTRODUCTION

There is a deep misunderstanding of what constitutes victory in cyber conflict. In most writing on cyber power, what constitutes 'winning' or 'losing' is rarely ever specified. When winning is even discussed, it is most often applied to actions on the network, whether a target is taken down or intruded into; the basic units of analysis are computer systems or malicious ones and zeroes. Because cyber conflict is seen as such a technical area, this tactical and technical view has been the dominant view for decades, perhaps ever since 1991 when the idea of a 'digital Pearl Harbor' first took root.

However, this view gives far too much attention to actions on the network and not enough to actual national security outcomes. This is most apparent in the Estonian cyber conflict of 2007, widely seen in the US military as a defeat in which Estonia was 'wiped off the net,' but which was in fact a fairly crushing operational and strategic win for the Estonians.

Since cyber power is becoming far too important for such a fundamental misunderstanding, where a victory is mistaken for a defeat, this paper analyses past writing on the topic, distinguishing efforts which focus on the tactical or technical level from those at the strategic level. The paper examines the two most illustrative case studies: Russian patriotic hackers against Estonia in 2007, and the long back-and-forth campaigns between the United States / Israel and Iran.

Defining victory in cyberspace, the next section argues, has been difficult for several reasons, including those mentioned above. However winning itself ought to be described as like any other kind of national security endeavour; that is, leading to better national security outcomes. Those outcomes might, in cyberspace, come from maximising hard power (espionage and attack capabilities), soft power (alliances, private-sector partnerships, influence), or economic power (trusted cyber companies, and a strong Internet-enabled economy). The paper then concludes with recommendations for policy-makers.

2. PAST WRITING ON WINNING AND LOSING

Winning is like 'power', in that it is 'surprisingly elusive and difficult to measure. But such problems do not make a concept meaningless.' (Nye, 2010). Yet to date, very little writing on winning or losing in cyber conflicts has been particularly specific about just what either might mean. There has been a strong tendency, especially in news articles or opinion pieces, to insist that the 'other guys' (often China, at least in the United States) are winning and 'our side' is

doomed to lose until the writer's recommendations are implemented. The most useful literature (including some with these weaknesses) falls into two relatively neat categories: the tactical and technical or operational levels, and the broader strategic or national security.

A. Winning and losing at the technical and tactical or operational level

The first set of useful literature focuses particularly on 'getting hacked' (if thought of as a technical issue) or on particular engagements or strings of engagements as part of a campaign (if thought of, in military terms, as the tactical and operational levels of war). Many pieces in this type imply that winning means keeping at bay criminal hackers, such as a recent corporate report on *Winning the Cyber War*. This report, published by Deloitte in cooperation with Symantec, recommends 'a clearer understanding of the risks posed to them and the level of protection needed to combat these threats, in order to inform an effective data protection strategy' (Deloitte, 2015).

More serious debate on war as it is meant in national security terms centres on winning as a successful use of cyber capabilities in a national security-related engagement or campaign.

The best example here is not of winning, but of losing, and not just any cyber battle, but a 'digital Pearl Harbor', where '[a]n aggressor nation could ... derail passenger trains, or even more dangerous, derail passenger trains loaded with lethal chemicals', as then-Defense Secretary Leon Panetta expressed it in late 2012 (Bumiller and Shanker, 2012). While the concept of an 'electronic Pearl Harbor' has been around (and much derided) since 1991, it does have an implied meaning of victory and defeat: the attacker succeeds in a devastating first strike against military or civilian targets.[1]

The attacker wins if their capabilities work as expected and take down their targets; in the zero-sum world of employment of cyber capabilities, the defenders have lost. This use of 'winning' appears to be precisely what Panetta had in mind, as immediately after his speech discussing a digital Pearl Harbor:

> 'Defense officials insisted that Mr. Panetta's words were not hyperbole, and that he was responding to a recent wave of cyberattacks on large American financial institutions [later blamed on Iran]. He also cited an attack in August on the state oil company Saudi Aramco, which infected and made useless more than 30,000 computers' (Bumiller and Shanker, 2012).

General Keith Alexander (former Director of the National Security Agency and Commander of US Cyber Command) highlighted the military view of winning engagements, saying that 'leaders want to dominate cyber-space in any encounters with adversaries' (Alexander, 2014). Most US military services have organisations committed to 'information dominance', such as the Army Information Dominance Centre and the Navy Information Dominance Corps. The top Air Force information officer also wears the hat of 'Chief of Information Dominance'.

It is probably this view that led Alexander, and his successor Admiral Michael Rogers, to see the attack on Sony Motion Pictures as a win for North Korea. The attacks destroyed data and

1 The first known use was by Winn Schwartau in congressional testimony on 27 June 1991.

cost perhaps hundreds of millions to clean up. Alexander later emphasised that 'destructive attacks that destroy data permanently, whether it hits Sony, a power [grid], a government, financial institution, the destruction of data in cyberspace is bad', and by implication a defeat as well (Gonzalez, 2014).

In the same vein, the hacker collective Anonymous has been said to be winning the war on Islamic State because it is taking down websites or social media accounts (Roberts, 2015), and the number of websites taken down has been a tally for success in cyber conflicts from Israel-Palestine in 2000 (Gambill, 2000), to India-Pakistan in 2000 (Nanjappa, 2010) and 2015 (Singh, 2015).

There are also classic examples from the US and Israeli cyber conflict with Iran as well as the 2007 attacks on Estonia, which will be examined in more depth below.

B. Winning and losing at the strategic level

The second set of useful literature looks at winning at the strategic level (the analysis of national power, far grander than tactics or operations), which is not the same as succeeding with a strategic attack (such as massive bombing to weaken an enemy's morale or ability to wage war). Winning at the tactical, technical, or operational levels means keeping out hackers or successful engagements; winning at the strategic level has far more at stake: strategic political objectives, broad national advantage or military defeat.

A first key theme is that defeat in a cyber battle might be defined best by the impact on a traditional military fight; for example, that successful intrusions give 'China the potential capability to delay or disrupt U.S. forces without physically engaging them – and in ways it lacks the capability to do conventionally' (US China Commission, 2008).

China is often portrayed as the victor in cyber espionage against the United States (and other countries). China is seen as winning because repeatedly successful campaigns over time are seen as giving it the upper hand in national security competition, even dominance costing billions of dollars in annual harm (US China Commission, 2015; IP Theft Commission, 2013). Even more starkly, Russia is believed not to have had to unleash a 'conventional' cyber conflict of denial of service attacks in the Estonia 2007 model because it already had such cyber dominance that such attacks would be counterproductive (Tucker, 2014). And even before the Snowden revelations, a common theme in my conversations with Chinese government officials was that everywhere they looked they saw the United States on the commanding heights of cyberspace.

Kuehl and Miller (2009) helpfully examined how losing at the tactical level could force a loss at the strategic level. America tends to lose such first battles, but then rally and win the follow-on war. But perhaps US dependence on cyber infrastructure would lead to a different result if America lost the first cyber battle in a digital Pearl Harbor? They refer to perhaps the classic definition of winning, but one that shows up surprisingly rarely in the cyber literature: '[w]inning that future war – defined in Clausewitzian terms as the attainment of strategic political objectives – thus may depend on successfully waging and winning the 'first battle in cyberspace'.

The Stuxnet attack against Iranian nuclear enrichment capability is one of the few occasions where the cyber discussion has focused on winning as achieving strategic political objectives. This attack disabled up to 1,000 Iranian centrifuges (Sanger, 2012), but did this meet the objectives? If the goal was to stop the programme, then it was a failure, as it did not succeed in stopping Iran or denting the regime's determination to develop a nuclear weapons capability (Rid, 2013; Barzashka, 2013; Lindsay, 2013). However, if the goal was to impose even a slight delay and keep the Israelis from launching pre-emptive air strikes, then perhaps the White House can feel that it won. This view aligns with definitions of cyber power such as 'the ability to use cyberspace to create advantages and influence events in other environments and across the instruments of power' (Kuehl, 2009).

One reason why successfully meeting political objectives seems to be such a scarce metric for winning is that it is hard to find any evidence of any strategically successful cyber attacks (Healey, 2013; Valeriano & Maness, 2015; Gartzke, 2013). According to Rid, 'the political instrumentality of cyber attacks has been starkly limited and is likely to remain starkly limited'.

Indeed, even when 'winning' might seem straightforward, as in the case of Chinese espionage, the analysis is almost entirely about what the United States has lost, not how the stolen intellectual property has enriched China. The United States may indeed be having hundreds of billions of dollars' worth of intellectual property stolen, but how much of that gain is China actually capturing, and is it enough to meet Chinese political objectives, given the price they have paid in worsening relations with nations from whom it has stolen?

3. CASE STUDIES OF MAJOR CYBER CONFLICTS

Many of the specific case studies of cyber conflict remain classified by national intelligence services or hidden by corporations not willing to release data for fear of embarrassment or because of its commercial value. Accordingly, researchers are forced to use a relatively limited set of cases from which to draw conclusions. Fortunately, two of these, Estonia and the conflict between Iran and the United States and Israel, are both illustrative and important case studies. There is a wealth of public information, and both have been discussed frequently by key policy-makers.

A. Estonia, 2007

The 2007 campaign of cyber attacks against Estonia by Russian patriotic hackers, ignored or even encouraged by the Kremlin, is by far the most instructive. In the context of rising Estonian identity to the perceived detriment to Russian interests, the proximate cause of the campaign was the Estonian government choosing to move the Bronze Soldier, a statue of a Red Army soldier, along with the bodies of several unknown soldiers from World War Two. The relocation, from the centre of Tallinn to a military cemetery, led to a couple of nights of rioting in the city by Russian-speaking Estonians, and a campaign of denial of service attacks in several waves over the next few weeks, peaking on 9 May, the anniversary of the defeat of Nazi Germany (Schmidt, 2013). It was a classic cyber conflict, but not a cyber war; no one died from the cyber attack, and it caused only transient harm to disabled government, financial and other websites.

The immediate goals of the attacks appear to be straightforward; to use cyber capabilities as a 'cyber riot' or protest, to express displeasure over the move and to coerce the government to cancel it. This was a common message of the Russian-language media at the time. In the broader context, the goal was probably to help flex Russian power in the perceived 'near abroad' state of Estonia, once part of the mighty Soviet Union but now part of NATO. By ignoring or encouraging cyber attacks by Russian patriotic hackers, including condemnation by Russian President Vladimir Putin against those who 'desecrate memorials to war heroes' (BBC News, 2007), Moscow could send a message without seeming to be directly involved (Healey, 2013).

Most analyses of the campaign pivot around the theme that the Russian attack was so successful that Estonia 'more or less cut itself off from the internet' (Economist, 2010), as I have had General Alexander tell me (Alexander, 2013). Other Pentagon cyber officials were even more pointed: 'Why are the Estonians considered cyber experts? All they did is get knocked off the net'. Many Estonian websites were indeed knocked offline, and the Russian patriotic hackers were flooding Estonian networks with so much traffic that the Estonians were forced to disconnect themselves from international traffic at their local Internet Exchange Point (Schmidt, 2013). But, unfortunately, this analysis of the tactical and technical truths of the campaign is also beside the larger point. The conflict might have been a tactical defeat for the Estonians, but it was a clear operational win. Even after several weeks of disruptive attacks, the Estonians still moved the statue. That is, they refused to be coerced by the Russian exercise in cyber power. One of the smallest nations on the planet, with just 1.3 million people, no tanks, and no fighter aircraft, stood up to one of the most powerful.

Indeed, if the Kremlin's larger goal in ignoring and encouraging the attacks was to flex Russian power and keep Estonia cowed, then the campaign was a strategic loss for Russia as well. The Kremlin, after the campaign of attacks, emerged with its power somewhat diminished. NATO considered the Russian cyber attack on a member as a 'wake-up call' for the alliance, leading to response plans, wargames, and the creation of a new cyber centre of excellence – in Tallinn, Estonia (Healey and Jordan, 2014). The Estonians, in part because of the 2007 attacks, were feted in Washington and European capitals and are 'renowned for cyber security' (Economist, 2012).

The Estonians did not lose in 2007; in fact it wasn't even close. This means that the US military, the military force that has most embraced cyber capabilities, cannot distinguish between a loss and a win. And if the military cannot distinguish between those, then it will have a seriously unbalanced understanding of cyber power.

B. US – Iran, 2008 to 2015

Unfortunately, not all analyses of cyber conflict are as straightforward as that of Estonia. The ongoing cyber conflict between Iran and the United States has not just been a single campaign, but a years-long back and forth, with each side apparently giving as good as they've got.

Seemingly the first shot in the cyber conflict, at least from public sources, was the US-Israeli Stuxnet attack against Iranian uranium enrichment capability (Sanger, 2012). This first-ever truly destructive cyber attack, conducted from around 2008 onwards, was certainly a tactical

and operational success, destroying at least 1,000 centrifuges (Zetter, 2014). Whether or not it was a strategic win is harder to gauge, as Iranian enrichment actually increased during this period, though the cyber attack did perhaps help forestall an Israeli air strike.

Unfortunately, Stuxnet was an operational success that forced the Iranians to redouble their own cyber capabilities, which they had generally ignored until attacked (Fox-Brewster, 2014). In 2012, Iran appears to have conducted a massive and sustained denial-of-service attack against many US banks, including Bank of America, Citigroup, Wells Fargo, PNC and others, 'most likely in retaliation for economic sanctions and online attacks by the United States' (Perlroth and Hardy, 2013). The attacks were 'unprecedented' in 'scale, scope, and effectiveness' and seen as 'the first significant digital assault of its kind undertaken against American industry's computers by a foreign adversary' (Nakashima, 2014).

As the attacks were reported as 'major disruptions' and took place over months, they were probably considered a tactical and technical success; if the ultimate political purpose was to disable the entire finance sector or to coerce US policy towards Iran, then they fell far short. If, instead, the goal was retribution for sanctions or to attract the attention of US policy-makers, then perhaps the Iranians considered that they won that round.

At about the same time, another series of tit-for-tat attacks were taking place. Iran was forced, in April 2012, to disconnect some oil wells from the Internet in the face of the damaging Wiper worm, which wiped the hard drives of computers of several oil companies and bureaucracies (Erdbrink, 2012). According to most analyses, the attack was most likely the work of Israel, keen to disrupt the Iranian economy (Zetter, 2014). Just a few months later, in August, a nearly identical wiper attack, called Shamoon, took down nearly 30,000 computers at Saudi Aramco, and more a few days later at RasGas (Perlroth, 2012). A leaked NSA document meant for the NSA director, General Keith Alexander, 'suggests that the attack on Saudi Aramco was in response to 'a similar cyberattack' against Iran's oil industry earlier that year' (Sanger, 2015).

Shamoon has been called a 'game changer' even more significant than Stuxnet, as it was so destructive. But if the Iranian goal was to disrupt Saudi oil and Qatari gas production, then the attack was an operational and strategic flop. There was no disruption of production. American officials who point to Shamoon as a game changer might be technically correct, but are highlighting an attack which was not just a somewhat symmetrical retaliation but which largely failed.

On balance, the individual engagements – Stuxnet, bank DDoS attacks, Wiper, and Shamoon – were all tactical wins. Other than Stuxnet, most seem to have been operational defeats, or at least failures, as the damage inflicted was neither severe nor lasting enough. Not knowing what the political objectives were for Israel, Iran, or the United States, it is even harder to identify any strategic victor.

As a footnote, in the less combative atmosphere after the nuclear agreement with Western powers in 2015, Iran has shifted its cyber operations away from such destructive attacks to espionage (Bennett, 2015).

4. ANALYSIS

As in many other areas of warfare, tactical and technical gains seem relatively easy but translating those into larger political and military success is far harder. It is likewise easiest to determine the victor at the more tactical and technical levels.

A. Difficulty of assessing who won

At the strategic level, the question of 'who won' defies an easy binary answer, for many reasons: an over-focus on the tactical and technical levels of cyber operations; a military fascination with capabilities, especially usable ones; the nature of irregular conflicts; a lack of a national cyber strategy; and high security classifications.

Cyber conflict is also fought over networks using bits and bytes by cyber warriors who are steeped in technical know-how, so it is no surprise that winning or losing have been such a technical determination. Cyber conflict is indeed fast at the tactical level; ones and zeroes do indeed travel at network speed. But the public evidence is clear that individual engagements have rarely if ever made a strategic, national security difference (Lindsay, 2013; Healey, 2013).

Tied to this technical mind-set is a strong military focus, not on operations, doctrine, or strategy but on capabilities. Any time a military officer or leader discusses cyber capabilities, they are in some way reinforcing a tactical and technical mentality, fighting from the foxhole or an endless series of dogfights rather than with a larger picture. I have been in meetings with the most senior Air Force cyber officers where the discussion never got farther than what capabilities will set the Air Force apart from other services, rather than what cyber conflict of the future might be like and how the United States and its Air Force can prevail.

Cyber capabilities are seen as usable in an era when more overt military power is largely forbidden, so 'doing something' in a post-Vietnam, post-9/11 era becomes almost a political end in itself. Along with special forces (whether SEAL trainers or "little green men"), proxy non-state groups, or drone strikes, nations engage in offensive or espionage cyber operations for relatively limited tactical gains, divorced from longer-term strategic outcomes. Likewise, cyber conflicts tend to be irregular conflicts, and it is almost always difficult to determine winning or losing in such fuzzy and indistinct circumstances. Counter-terrorism experts voice the same frustrations on whether their successful operations are leading to an overall win or a loss. Likewise, determining if the United States or Iran won their cyber engagements is in some ways no more or less difficult than deciding if the United States is winning in Iraq, Afghanistan, or Syria. The goals of those conflicts have reduced significantly over the years, and winning and losing have been redefined many times.

This is especially difficult for the United States, as it does not have a single cyber strategy by which to assess victory or defeat. Rather, there are separate strategies for the military, commerce and trade, and international issues. The counterinsurgency warfare community has had a decades-long debate on whether winning was best pursued by winning the hearts and minds of the citizens or by killing the insurgents, so that tactics and operations could be

balanced against these larger goals. Lacking such a cyber strategy, the United States is hobbled in trying to answer critical questions such as whether risky attacks like Stuxnet are worth the danger. With so much post-9/11 power centred in the Pentagon, the military view of winning becomes the default.

In addition, nations ensure that their offensive and espionage cyber operations are highly classified. It is difficult for anyone other than a handful of extremely highly cleared military officers and senior officials to know which side is winning in the Iran versus US and Israel conflict. Worse, US cyber warriors classify or downplay any outgoing cyber attacks, then loudly denounce attacks or even counterattacks by the other side. The best examples involve Iran, where the US and Israel threw the first punch with Stuxnet. US officials like General Alexander later downplayed that attack, which nearly all other experts, including his predecessor as NSA Director, considered to be 'crossing the Rubicon' into a new era of cyber conflict (Sanger, 2012), to rather highlight the Iranian Shamoon attack (InfoSec Magazine, 2014). But Shamoon was in fact apparently a relatively proportionate retaliation to the earlier Wiper attack on Iran's own energy industry. Listening only to statements by General Alexander might lead experts to believe that the United States was losing to an aggressive Iran.

B. Toward a better understanding of winning

These reasons make it difficult to think clearly about winning or losing in cyber conflict, but there is still room for significant progress. Cyber warriors tend to see that Estonia lost in 2007 because of a focus on technical impact rather than the more strategic view that winning means achieving better national security outcomes. Estonia won because it emerged from 2007 far stronger and has spent the better part of a decade building on that strength; Russia came out weaker relative to Estonia, NATO, and other perceived adversaries.

Accordingly, the most important next step for many nations facing cyber conflict is to be far clearer about preferred national security outcomes; these should be prioritised in a clear national cyber strategy to clarify decisions when a government is faced with competing public goals.

Those national security outcomes might define winning in different ways:

- **Hard-power perspective:** Winning in cyberspace is dominating it for espionage and offensive operations, with the best arsenal of cyber capabilities and 'collecting the whole haystack' of worldwide ones and zeros, as one official described the NSA approach (Nakashima & Warrick, 2013).
- **Soft-power perspective:** Winning in cyberspace is to seize this once-in-a-century opportunity to win the hearts and minds of the digital natives in whatever nation they live, so that they see our nation as representing their values and enriching their lives, giving concomitant influence.
- **Economic-power perspective:** 'Winning in cyberspace' is to have the most agile Internet-enabled economy, the strongest technology companies, and the most trusted cyber security experts.

If the militarised view of long-term national security outcomes turns out to be the historically correct perspective, then nations like the United States and Russia are on a strong path. If, however, either the soft-power or economic perspectives are likely to lead to better outcomes, then the United States and nations that copy its views on cyber power are headed for potentially far worse outcomes.

Nations which follow the hard-power perspective are likely to win battles, but never the war.

5. RECOMMENDATIONS

Cyber espionage and attack are ultimately perhaps too ingrained as modern tools of statecraft to see a drastic reduction, especially to other forms of irregular conflict. For example, if a head of government wanted to ban assassinations (as President Gerald Ford did in 1976) or stop using drones for targeted killing of terrorists, the decision and execution are relatively straightforward; they have but to say that these actions are not in the long-term interest of the nation, and direct that they stop. With nearly the whole world being wired, it is now the 'golden age of espionage' for intelligence services using cyber means, while militaries see adversaries increasingly adding networking capabilities to ever more juicy-looking targets. Few heads of government could simply demand that it all stop. Accordingly, solutions are more about the degree of emphasis and process.

To align around a better view of using cyber power to win, nations – starting with the United States – need to take the following actions:

1. **Create an overarching cyber strategy** that is clear about which long-term national cyber priority is the most important: hard power, soft power, or economic power. Strategies cannot, as has often been the norm in the United States and other nations, list multiple competing priorities pursuing often competing public goods. This document should be driven, and signed, by the head of government. If a full 'strategy' is too difficult for bureaucratic or other reasons, just a clearly delivered policy preference by the head of government can be sufficient.
2. **Revamp the interagency policy process** to deliver on the priority chosen to deliver those long-term best outcomes. For example, major military or espionage campaigns cannot be shielded from scrutiny for classification reasons, or approved by only a narrow base of cleared individuals often with little experience or concern of non-military or -intelligence matters.
3. **Encourage a broader understanding of cyber power**, including how future cyber conflict might differ from what is seen today; the interplay of soft power and economic power on the results of cyber conflict; the role of the private sector in delivering victory; and the differences between the tactical, operational and strategic levels of cyber conflict. With the current mind-set so ingrained in many governments and militaries, this broader dialogue probably needs to be led by academia and think tanks, perhaps supported by grants.

4. **Promulgate broader thinking in subordinate strategies and doctrine**. The view of 'victory' that is decided on by governments needs to trickle down into the bureaucracy, especially into individual ministries' cyber strategies and projects, military strategy and doctrine, and into military academies so that the next generations of military practitioners and leaders learn the best way to do things, not the merely the past and current way.

It is apparent that there is a deep misunderstanding of what constitutes victory in cyber conflict, with far too much attention on actions on the network and not on actual national security outcomes. This is most apparent in the Estonia cyber conflict of 2007, widely seen in the US military as a defeat, but which was in fact a fairly crushing operational and strategic win for the Estonians.

Cyber power is becoming far too important for such a fundamental misunderstanding, where a victory is mistaken for a defeat. Fortunately, there is a relatively straightforward set of recommendations, starting with a clear national priority set by the head of government, which clearly points to a clearer path.

REFERENCES

Alexander, Gen Keith, interview by Christopher Joye. 2014. 'Interview transcript: former head of the NSA and commander of the US cyber command, General Keith Alexander' *Financial Review*, May 9. http://www.afr.com/technology/web/security/interview-transcriptformer-head-of-the-nsa-and-commander-of-the-us-cyber-command-general-keith-alexander-20140507-itzhw.

Alexander, Keith, interview by Jason Healey. 2013. (May 3).

Barzashka, Ivanka. 2013. 'Are Cyber-Weapons Effective?' *RUSI Journal* (Royal United Services Institute) 158 (2).

BBC News. 2007. 'Putin in veiled attack on Estonia' May 9. http://news.bbc.co.uk/2/hi/europe/6638029.stm.

Bennett, Cory. 2015. 'Iran launches cyber offensive after nuclear deal'. *TheHill.com* November 24. http://thehill.com/policy/cybersecurity/261190-iran-switches-to-cyber-espionage-after-nuclear-deal.

Bumiller, Elisabeth, and Tom Shanker. 2012. 'Panetta Warns of Dire Threat of Cyberattack on U.S.' *New York Times* October 11. http://www.nytimes.com/2012/10/12/world/panetta-warns-of-dire-threat-of-cyberattack.html?_r=0.

Deloitte. 2015. *Winning the Cyber War*. London: Deloitte. http://www2.deloitte.com/content/dam/Deloitte/uk/Documents/consultancy/deloitte-uk-symantec-deloitte-partnership-uk.pdf.

Economist. 2010. 'War in the fifth domain.' *The Economist*, June 1. http://www.economist.com/node/16478792.

Economist. 2012. 'Paper Cuts.' *The Economist*. October 27. http://www.economist.com/news/international/21565146-paperless-polling-stations-are-unfashionable-internet-voting-its-way-paper-cuts.

Erdbrink, Thomas. 2012. 'Facing Cyberattack, Iranian Officials Disconnect Some Oil Terminals From Internet.' *New York Times*. April 23. http://www.nytimes.com/2012/04/24/world/middleeast/iranian-oil-sites-go-offline-amid-cyberattack.html?_r=0.

Fox-Brewster, Thomas. 2014. '"Bone-Chilling" Research Suggests Iran Gearing Up To Avenge Stuxnet Hacks.' *Forbes.com* December 2. http://www.forbes.com/sites/thomasbrewster/2014/12/02/bone-chilling-research-suggests-iran-gearing-up-to-avenge-stuxnet-hacks/.

Gartzke, Eric. 2013. 'The Myth of Cyberwar.' *International Security* 41-73. http://www.mitpressjournals.org/doi/pdf/10.1162/ISEC_a_00136.

Gambill, Gary. 2000. 'Who's Winning the Arab-Israeli Cyber War?' November. https://www.meforum.org/meib/articles/0011_me2.htm.

Gonzalez, Eileen. 2014. 'Retired General: Sony cyber attack is an act of war'. December 2. http://www.ksat.com/news/retired-general-sony-cyber-attack-is-an-act-of-war.

Healey, Jason, ed. 2013. *A Fierce Domain: Conflict in Cyberspace, 1986 to 2012*. Cyber Conflict Studies Association.

Healey, Jason. 2013. 'Concluding Assessment.' In *A Fierce Domain: Conflict in Cyberspace, 1986 to 2012*, by Jason Healey. Cyber Conflict Studies Association.

Healey, Jason, and Klara Tothova Jordan. 2014. *NATO's Cyber Capabilities: Yesterday, Today and Tomorrow*. Issue brief, Atlantic Council. http://www.atlanticcouncil.org/images/publications/NATOs_Cyber_Capabilities.pdf.

Infosecurity Magazine. 2014. 'Saudi Aramco Cyber Attacks a 'wake-up call', Says Former NSA Boss'. May 8. http://www.infosecurity-magazine.com/news/saudi-aramco-cyber-attacks-a-wake-up-call-says/.

IP Theft Commission. 2013. *Report on the Commission on the Theft of American Intellectual Property*. National Bureau of Asian Research. http://www.ipcommission.org/report/ip_commission_report_052213.pdf.

Kay, Kim. 2015. *How to Win the Cyber War*. November 30. http://wwpi.com/how-to-win-the-cyberwar/.

Khanna, Parag. 2015. 'How small states prepare for cyber-war'. *CNN.com* September 2. http://www.cnn.com/2015/09/02/opinions/estonia-cyber-war/.

Kuehl, Dan, and Robert A Miller. 2009. *Cyberspace and the 'First Battle' in 21st Century War*. Washington DC: National Defense University Press. http://ndupress.ndu.edu/Portals/68/Documents/defensehorizon/DH-68.pdf.

Kuehl, Daniel. 2009. 'From cyberspace to cyberpower: Defining the problem.' In *Cyberpower and National Security*, edited by Franklin Kramer, Stuart Starr and Larry Wentz. National Defense University.

Lindsay, Jon. 2013. 'Stuxnet and the Limits of Cyber Warfare.' *Security Studies*. http://www.tandfonline.com/doi/pdf/10.1080/09636412.2013.816122.

Mite, Valentinas. 2007. 'Estonia: Attacks seen as first case of "cyberwar"'. May 30. http://www.rferl.org/content/article/1076805.html.

Nakashima, Ellen. 2014. 'U.S. rallied multinational response to 2012 cyberattack on American banks'. *Washington Post* April 11. https://www.washingtonpost.com/world/national-security/us-rallied-multi-nation-response-to-2012-cyberattack-on-american-banks/2014/04/11/7c1fbb12-b45c-11e3-8cb6-284052554d74_story.html.

Nakashima, Ellen, and Joby Warrick. 2013. 'For NSA chief, terrorist threat drives passion to "collect it all"'. *Washington Post* July 14. https://www.washingtonpost.com/world/national-security/for-nsa-chief-terrorist-threat-drives-passion-to-collect-it-all/2013/07/14/3d26ef80-ea49-11e2-a301-ea5a8116d211_story.html.

Nanjappa, Vicky. 2010. 'Cyber wars: Pak has an advantage over India'. *Rediff.com* August 16. http://www.rediff.com/news/report/indo-pak-cyber-war-set-to-escalate/20100816.htm.

Nye, Joseph. 2010. *Cyber Power*. Harvard Kennedy Schook, Belfer Center. http://belfercenter.ksg.harvard.edu/files/cyber-power.pdf.

Perlroth, Nicole. 2012. 'In Cyberattack on Saudi Firms, U.S. Sees Iran Firing Back'. *New York Times* October 23. http://www.nytimes.com/2012/10/24/business/global/cyberattack-on-saudi-oil-firm-disquiets-us.html.

Perlroth, Nicole, and Quentin Hardy. 2013. 'Bank Hacking Was the Work of Iranians, Officials Say'. *New York Times* January 8. http://www.nytimes.com/2013/01/09/technology/online-banking-attacks-were-work-of-iran-us-officials-say.html?_r=1.

Rattray, Gregory. 2001. *Strategic Warfare in Cyberspace*. MIT Press.

Rid, Thomas. 2013. *Cyber War Will Not Take Place*. Oxford University Press.

Roberts, Thomas. 2015. 'Is Anonymous' cyber war against ISIS working?' November 23. http://www.msnbc.com/thomas-roberts/watch/is-anonymous-cyber-war-against-isis-working-572660291779.

Sanger, David. 2015. 'Document Reveals Growth of Cyberwarfare Between the U.S. and Iran'. *New York Times* February 22.

Sanger, David. 2012. 'Obama Order Sped Up Wave of Cyberattacks Against Iran'. *New York Times* June 1. http://www.nytimes.com/2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html.

Schmidt, Andreas. 2013. 'The Estonian Cyberattacks.' In *A Fierce Domain: Cyber Conflict, 1986 to 2012*, by Jason Healey. Cyber Conflict Studies Association.

Singh, Ritu. 2015. 'Cyber-war: Indian hackers hack 250+ Pakistani websites after attack on Kerala govt's website'. ZeeNews September 29. http://zeenews.india.com/news/net-news/cyber-war-indian-hackers-hack-250-pakistani-websites-after-attack-on-kerala-govts-website_1803606.html.

Tucker, Patrick. 2014. 'Why Ukraine Has Already Lost The Cyberwar, Too'. *Defense One* April 28. http://www.defenseone.com/technology/2014/04/why-ukraine-has-already-lost-cyberwar-too/83350/.

US China Commission. 2008. *US China Commission 2008 Annual Report*. US GPO. http://origin.www.uscc.gov/sites/default/files/annual_reports/2008-Report-to-Congress-_0.pdf.

US China Commission. 2015. *US China Commission 2015 Annual Report*. US GPO. http://origin.www.uscc.gov/sites/default/files/annual_reports/2015%20Annual%20Report%20to%20Congress.PDF.

US Department of Defense. 2013. *Joint Publication 3-12, Cyberspace Operations*. http://www.dtic.mil/doctrine/new_pubs/jp3_12R.pdf.

Valeriano, Brandon, and Ryan Maness. 2015. *Cyber War Versus Cyber Realities: Cyber Conflict in the International System*. Oxford University Press.

Zetter, Kim. 2014. *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon*. Crown.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Towards a Theory of Cyber Power: The Israeli Experience with Innovation and Strategy

Lior Tabansky
Blavatnik Interdisciplinary
Cyber Research Center (ICRC)
Tel Aviv University (TAU)
Tel Aviv, Israel
cyberacil@gmail.com

Abstract: Cyber power has become a topical issue for many democracies, but policy and scholarly debates often default to the cyber means alone. However, throughout history, superior means were never enough to secure strategic advantage. Strategy – seeking various ways and diverse means to serve clear ends – is the missing ingredient.

I outline indicators of cyber power and develop an interdisciplinary framework for strategic analysis. Cyber power manifests itself when one leverages means to operate in cyberspace towards achieving a political end. I apply this strategic ends-ways-means framework to an Israeli case study to determine its scholarly value. The analysis suggests that Israel demonstrated cyber power when applying various means towards achieving ends beyond enhanced cyber security. In soft power efforts, Israel harnessed cyber technology for economic growth and increased cooperation with like-minded nations. Israel purportedly developed and applied cyber warfare to attain its top strategic priority through hard power – preventing Iran from acquiring nuclear weapons. Another finding challenges conventional wisdom on the value of formal policies for national cyber power.

Security studies scholarship on strategy and economics scholarship on National Innovation Systems facilitate improved understanding of soft and hard cyber power. Strategic studies offer valuable insights into adaptation process, which can help policy makers avoid predictable pitfalls. Fostering society-wide innovation capacity crucially helps to better adapt to the volatile future. The National Innovation System scholarship helps to comprehend and obtain better means. Scholars of cyber power should venture further beyond the core technical disciplines.

Keywords: *strategy, Israel, Stuxnet, National Innovation System, R&D*

1. STRATEGY: THE MISSING INGREDIENT OF CYBER POWER

Cyber technology provides new and affordable tools for actors to pursue their interests.[1] Unsurprisingly, cyber debates often default to a focus on the more easily quantifiable technology: networking and system architecture, cryptography, malware samples, military commands, and cyber defender headcounts. Despite years of effort and many billions of dollars invested in vastly improved technology, cyber power remains elusive.

Western cyber insecurity is a familiar situation for a strategist: throughout history, superior *means* were never enough to secure strategic advantage. Cyber power manifests when one leverages *means* to operate in, or to mould, the man-made cyber substrate[2] towards achieving a political *end*. But in the recent words of a leading strategist,

> 'Senior people in the ranks of strategic studies have by and large ignored the growing cyber challenge, while those who are technically highly cyber knowledgeable typically have scant background in strategy'.[3]

Strategy – seeking various ways and diverse *means* to serve clear *ends* – is the missing ingredient in cyber power scholarship and policy. In this essay I develop an interdisciplinary analytical framework of cyber power, to bridge the gap between cyber technology and strategy. It stems from an ongoing interdisciplinary analytical effort to advance a more comprehensive understanding of cyber power in strategic studies and international relations, and builds on the author's first case study of Israeli cyber security policy.[4]

2. OUTLINE

This study uses strategic studies scholarship on strategy together with economics studies of National Innovation Systems to lay out the new interdisciplinary analytical framework for cyber power. Case studies in social science can be a source of new theoretical development and a powerful tool for testing theories. Theory-building from case studies is an increasingly popular and relevant research strategy that forms the basis of influential studies.[5] Qualitative research enables the researcher to capture the complexity of the object of study.[6] Thus, the Israeli empirical case study is analysed to demonstrate the way in which an interdisciplinary

1 Lucas Kello, 'The Meaning of the Cyber Revolution: Perils to Theory and Statecraft,' *International Security* 38, no. 2 (2013).
2 Chris C. Demchak, *Wars of Disruption and Resilience Cybered Conflict, Power, and National Security* (Athens, Ga.; London: The University of Georgia Press, 2011).
3 Colin S. Gray, *Making Strategic Sense of Cyber Power : Why the Sky Is Not Falling*, ed. Strategic Studies Institute Army War College (2013).
4 Lior Tabansky and Isaac Ben-Israel, *Cybersecurity in Israel*, ed. Sandro Gaycken, Springerbriefs in Cybersecurity (Springer, 2015).
5 Kathleen M. Eisenhardt, 'Building Theories from Case Study Research,' *The Academy of Management Review* 14, no. 4 (1989).
6 John W. Creswell, *Qualitative Inquiry and Research Design: Choosing among Five Approaches* (Sage publications, 2012).

strategic analysis of cyber power seeks the *ends-ways-means* nexus. The case selection is driven by the high regard for Israeli cyber security globally.[7]

The article demonstrates the framework's realistic analytical value: applied to the Israeli case study it reveals national cyber power and helps assessing it. The article presents the thesis in section 2, followed by a brief introduction to key themes in strategic studies in section 3. The Israeli case study is then presented in section 4: the country's main economic and social indicators of innovation; policy efforts; and its cyber warfare experience. Bridging the interdisciplinary gap, I harness strategic and economic scholarship to analyse selected Israeli academic, business and defence sector contributions to cyber power. Section 5 examines how cyber *means* support grand strategy *ends*, through various instrument of hard and soft power. The study's findings challenge the common wisdom on formal policy's role in national cyber power. The article's principal academic value lies in applying grand strategy and economics study of national innovation systems to analyse national cyber power. Future research directions are offered in section 6.

3. ON STRATEGIC THOUGHT

Strategic studies became an interdisciplinary academic field studying international conflict and peace strategies after WWII. However, strategic thought has been crucial since the time of the ancient civilisations.[8] Despite, or perhaps because of, technological and social change, power and strategy have remained essential.

A. Power

Power, like many basic ideas, is a contested concept. It depends on context, perception, and anticipation, not just on the application of force. Power is both a tool and a goal in itself, in peace no less than in war.

Joseph Samuel Nye, Jr., one of the most influential international relations scholars and a former chairman of the US National Intelligence Council, distinguished hard and soft power along a spectrum from command to co-option in a seminal 1990 article.[9] Hard power behaviour relies on coercion and payment, while soft power uses the framing of agendas, attraction, or persuasion. Nye also discussed cyber power, masterfully including both physical and informational instruments, soft and hard power aspects, and ramifications within and beyond cyberspace.[10] Cyber power is not limited to information, but cuts across the other facets, elements and instruments of power, often referred to as Diplomatic, Informational, Military, and Economic (DIME). Cyber connects these elements in new ways to produce preferred outcomes within and outside cyber space. Kuehl's definition set out the central concepts for cyber power:

7 B. Grauman, 'Cyber-Security: The Vexed Question of Global Rules: An Independent Report on Cyber-Preparedness around the World,' ed. Security & Defence Agenda (SDA) and McAfee Inc. (Brussels: Security & Defence Agenda (SDA), 2012). 'Cyber-Boom or Cyber-Bubble? Internet Security Has Become a Bigger Export Earner Than Arms,' *The Economist*, Aug 1 2015.

8 Sun Tzu, *The Art of War* (Shambhala Publications, 2011). Thucydides, *History of the Peloponnesian War*, The Penguin Classics (Harmondsworth, Eng.: Penguin Books, 1972).

9 Joseph S. Nye, 'Soft Power,' *Foreign policy* (1990).

10 Joseph S. Nye, '*Cyber Power*.' Belfer Center for Science and International Affairs, Harvard Kennedy School, May 2010.

> '…the ability to use cyberspace to create advantages and influence events in other operational environments and across the instruments of power'[11]

The very inclusion of the desired outcomes in the definition entails that such ends should be defined, and should guide the development and application of power.

B. Strategy

The term 'strategy' originated in the context of conflicts between city-states in ancient Greece, and in addition to its continued military uses, has been adopted by business, governments, political campaigns, and more, becoming ubiquitous.[12] In analysing cyber power, I adopt Sir Lawrence Freedman's recent definition: 'Strategy is about getting more out of a situation than the starting balance of power would suggest. It is the art of creating power'.[13]

By definition, getting more out of a situation presents obvious difficulties. In his seminal article, *Why Strategy is Difficult*, Colin S. Gray discussed three major reasons why it is difficult to do strategy well:

- Its very nature, which endures through time and in all contexts;
- The multiplicity and sheer variety of sources of friction; and
- It is planned for contexts that have not occurred and might not occur; the future has not happened.[14]

Cyber technology can offer many benefits; it cannot cure the Thucydidean 'honour, fear and profit' trinity, the human causes of policy already clear 2,400 years ago.[15] Strategic history suggests that developed states, tasked with securing their respective societies, are in for extraordinary shocks and surprises.[16] Recent strategic developments such as the Arab uprisings, the rise of Daesh, and Russia's moves in Ukraine and Syria, prove that Clauzewitz's fog and friction concepts remain valid.[17]

C. Grand strategy

The essence of strategy remains designing an effective relationship between *ends*, *ways* and *means* in potentially competitive or adversarial dynamic relations. In international power, an *'end'* is a *political* objective defined by the state's leadership. *'Way'* is the selected form of

11 Daniel T. Kuehl, 'Cyberspace and Cyberpower,' in *Cyberpower and National Security*, ed. Franklin D. Kramer, Stuart H. Starr, and Larry K. Wentz (National Defense University Press: Potomac Books, 2009).

12 Key strategic studies scholarship issues are covered in the edited collection of most of the influential essays: Thomas G. Mahnken and Joseph A. Maiolo, *Strategic Studies: A Reader* (Milton Park, Abingdon, Oxon; New York: Routledge, 2008).

13 Lawrence Freedman, *Strategy: A History* (Oxford: Oxford University Press, 2013).

14 Colin S. Gray, *Strategy and History: Essays on Theory and Practice* (London: Routledge, 2006).

15 Robert G. Gilpin, 'The Richness of the Tradition of Political Realism,' *International Organisation* 38, no. 02 (1984); Steven Forde, 'International Realism and the Science of Politics: Thucydides, Machiavelli, and Neorealism,' *International Studies Quarterly* 39, no. 2 (1995).

16 Max Boot, *War Made New: Technology, Warfare, and the Course of History, 1500 to Today* (New York: Gotham Books, 2006); Ian Arthur Bremmer, *The J Curve a New Way to Understand Why Nations Rise and Fall* (New York: Simon & Schuster, 2006); Paul M. Kennedy, *The Rise and Fall of the Great Powers Economic Change and Military Conflict from 1500 to 2000* (New York: Random House, 1987); Edward N. Luttwak, *Strategy the Logic of War and Peace* (Cambridge, Mass: Belknap Press of Harvard University Press, 2003).

17 Carl von Clausewitz, *On War*, trans. Michael Howard and Peter Paret (Princeton, N.J.: Princeton University Press, 1976).

action, boiling down to a mix of soft and hard power. *'Means'* refers to the resources available: people, money, land, trade influence, weapons etc. Strategy has several levels.

The focus of this article is the highest level, known as grand strategy:

> 'Grand strategy is a coherent statement of the state's highest political ends to be pursued globally over the long term. Its proper function is to prioritise among different domestic and foreign policy choices and to coordinate, balance, and integrate all types of national means – including diplomatic, economic, technological, and military power – to achieve the articulated ends'.[18]

Ideally, the political echelon defines the national strategy from which the security strategy is derived.[19] Alas, in practice rarely is it clearly and formally articulated. However, Edward Luttwak, the leading scholar on the hierarchical approach to defining grand strategy, writes: 'All states have a grand strategy, whether they know it or not.'[20]

TABLE 1: THE LEVELS OF STRATEGY[21]

Level	Geographic Scale	Temporal Scope	Types of Ends	Types of Power (Means)
Grand Strategy	Global	Long term (decades)	Highest political ends	All (diplomatic, informational, military, economic)
Strategy	All theaters of war (and conflict)	Mid term (years)	Overall military victory	Military, informational, economic
Operations	One particular theater of war	Short term (weeks to months)	Campaign victory	Military, informational
Tactics	Battlefield	Very short term (minutes to days)	Achievement of tactical objectives	Military
Technology	Home front/ academia Industry	Variable time horizon	Competitive advantage over enemies	Technical expertise

D. Israel's enduring grand strategy

Zionist political ideology emerged with modern nationalism in 19th century Europe, seeking self-determination through the establishment of a Jewish democratic state in the Land of Israel and the ingathering of the remaining Jewish diaspora to it.[22] But the volatile geo-political

18 Ibid.

19 Thomas G. Mahnken, 'U. S. Strategic and Organisational Subcultures,' in *Strategic Culture and Weapons of Mass Destruction: Culturally Based Insights into Comparative National Security Policymaking*, ed. Jeannie L. Johnson, Kerry M. Kartchner, and Jeffrey A. Larsen (New York: Palgrave Macmillan US, 2009).

20 Edward N. Luttwak, *The Grand Strategy of the Byzantine Empire* (Cambridge, Mass.: Belknap Press of Harvard University Press, 2009).

21 William C. Martel, *Grand Strategy in Theory and Practice: The Need for an Effective American Foreign Policy* (2015), p.30.

22 Anita Shapira, *Israel: A History*, http://site.ebrary.com/id/10628397.

environment has made the task daunting.[23] The founding fathers of Israel designed a national security strategy with the following elements continually present:[24]

- Seek qualitative superiority (including investment in education, science and technology);
- Seek an alliance with a global superpower, and normal diplomatic and economic relations with all countries;
- Emphasise early warning intelligence to balance the total lack of strategic depth (including heavy investment in signal intelligence); and
- Seek an ultimate deterrent (including heavy early investment in nuclear research).[25]

The overarching strategy, applied with varying degrees of prudence and effectiveness, has served the nation well. The Israeli population has grown ten-fold since 1948, and the GDP per capita has increased three-fold since 1990.[26] Israel was accepted into the OECD in 2010, and now ranks 18th among 188 nations on the UN's Human Development Index.[27] Recent political science scholarship shows the real-world effects for international governance and soft power that such ranking systems have.[28]

The geopolitical predicament persists. The implosion of the 1916 colonial Sykes-Picot political order in the Middle East along sectarian lines and the rise of global Jihadist organisations present volatile security challenges for Israel.[29] While Israeli national leadership avoids publishing formal national strategy documents,[30] Israel has viewed the nuclear ambitions of the Islamic Republic of Iran as the top strategic threat for over two decades.[31]

4. MEANS AND WAYS OF ISRAELI CYBER POWER

Having outlined the strategic ends, I now turn to survey the means and ways, towards a strategic analysis of cyber power. Technical innovation is central for cyber security. Israel is perceived

23 Avi Shlaim, *The Iron Wall: Israel and the Arab World* (2014).

24 Yisrael Tal, 'National Security the Israeli Experience,' Praeger, http://ebooks.abc-clio.com/?isbn=9780313001635; Yehezkel Dror, *Israeli Statecraft: National Security Challenges and Responses*, vol. 15, Besa Studies in International Security (Milton Park, Abingdon, Oxon; New York: Routledge, 2011); Efraim Inbar, *Israel's National Security: Issues and Challenges since the Yom Kippur War*, vol. 49, Cass Series-Israeli History, Politics, and Society (London; New York: Routledge, 2008).

25 Uzi Eilam, *Eilam's Arc: How Israel Became a Military Technology Powerhouse* (Brighton; Portland, Or.: Sussex Academic Press, 2011).

26 World Development Indicators 2015, (2015), http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=948695.

27 United Nations Development Programme, *Human Development Report 2015* (United Nations, 2016).

28 Judith G. Kelley, and Beth A. Simmons. "Politics by Number: Indicators as Social Pressure in International Relations." *American Journal of Political Science 59, no. 1* (2015).

29 Eran Zohar, 'Israeli Military Intelligence's Understanding of the Security Environment in Light of the Arab Awakening,' *Defence Studies* 15, no. 3 (2015).

30 Such as the French *Le Livre Blanc sur la Défense et la Sécurité Nationale* or the American *Quadrennial Defense Review*.

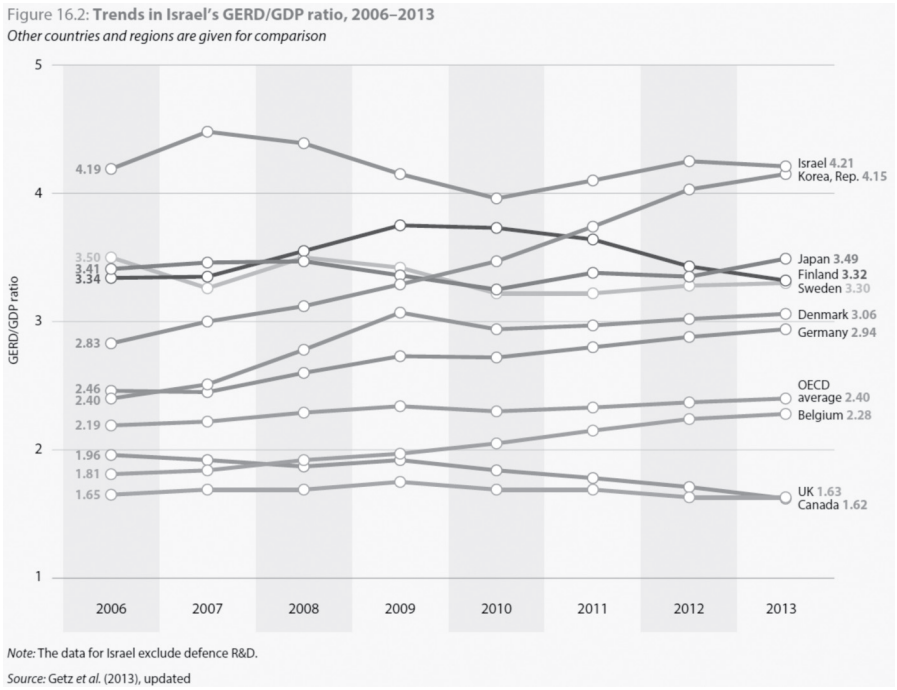
31 Ronen Bergman, *The Secret War with Iran: The 30-Year Clandestine Struggle against the World's Most Dangerous Terrorist Power* (Simon and Schuster, 2008); Wyn Q. Bowen and Jonathan Brewer, 'Iran's Nuclear Challenge: Nine Years and Counting,' *International Affairs* 87, no. 4 (2011); Yaakov Katz and Yoaz Hendel, *Israel Vs. Iran: The Shadow War* (Washington, D.C: Potomac, 2012). The Iranian nuclear program was first presented as an existential threat by Prime Minister Yitzhak Rabin in the early 1990s. In 2002, evidence of Iran's 'secret' nuclear program began to emerge. Israel's fear of an Iranian regime armed with a nuclear weapon takes at least three cumulative distinct forms: fear of annihilation, fear of a more difficult security environment, and fear of a challenge to Israel's founding Zionist ideological principles.

as a global leader in information technology.[32] Innovation capacity plays another, less tangible but important role; it indicates the likelihood of successful adaptation to change. The National Innovation System (NIS) concept refers to all the interacting social and political factors inside a country that affect the creation and diffusion of innovation. However, cyber capacity building debates have rarely used innovation studies, which have thrived in recent decades in economics, business management, political economy, technology, and engineering policy.[33]

A. The Israeli National Innovation System

Israel's gross domestic R&D expenditure is the highest in the world, and almost double the OECD average.

FIGURE 1: TRENDS IN ISRAEL'S GERD/GDP RATIO, 2006-2013[34]



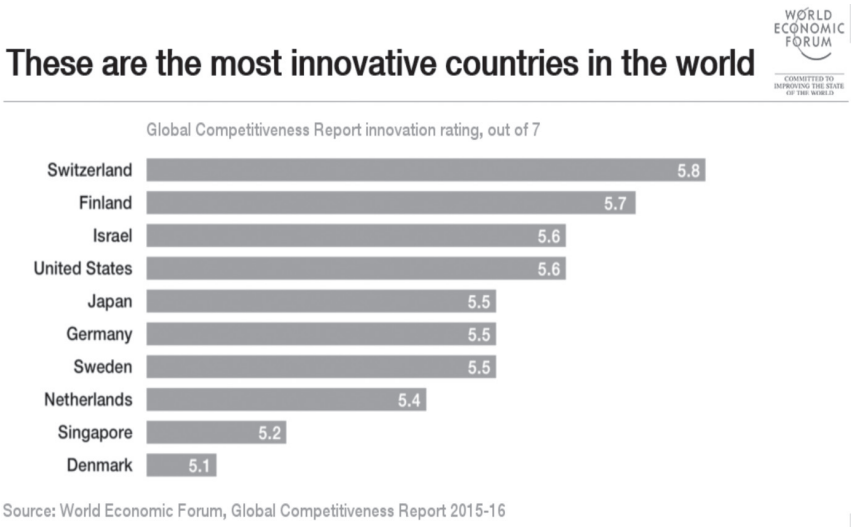
Importantly, the OECD figures exclude defence R&D expenditure. Israel ranks among the most innovative countries.

32 Grauman, 'Cyber-Security: The Vexed Question of Global Rules: An Independent Report on Cyber-Preparedness around the World; Dan Senor and Saul Singer, *Start-up Nation: The Story of Israel's Economic Miracl*e (New York: Twelve, 2009).

33 Mark Z. Taylor, 'Toward an International Relations Theory of National Innovation Rates,' *Security Studies* 21, no. 1 (2012).

34 OECD. 'R&D in OECD and Key Partner Countries, 2013.' Paris: OECD Publishing, 2015.

FIGURE 2: THE MOST INNOVATIVE COUNTRIES IN THE WORLD[35]



B. Universities

In the Israeli NIS, public research universities conduct basic research and participate in most applied research. Israeli universities compete globally, and have had remarkable success in the EU's Seventh Framework Programme (FP7).[36] Each university owns a technology transfer company (TTC) to protect and proactively commercialise scientific innovations made by researchers.

Regarding cyber security, Israeli universities host four of the top 50 Computer Science departments.[37] Tel Aviv University (TAU) hosts the Blavatnik Interdisciplinary Cyber Research Centre, the first institutionalised Israeli government-academia cooperative venture into cyber-related research. It was inaugurated in September 2014 by Prime Minister Netanyahu during TAU's 4th Annual Cyber security Conference.

C. Business R&D

Since the domestic market is small, Israel's industry can only prosper through exports. To succeed in global competition, the industry has to seek rich diversity and cutting-edge competitiveness. Even the Israeli defence industries export some 70% to 76% of the output.[38,39] This global orientation is one of the reasons that Business Expenditure on R&D (BERD) in Israel, as a share of GDP, is around 80%, the second highest in the OECD.[40] Half of Israel's

35 Klaus Schwab, 'Global Competitiveness Report, 2015-2016,' in *World Economic Forum* (2015).
36 Commission European, Research Directorate-General for, and Innovation, *Research and Innovation Performance in the EU: Innovation Union Progress at Country Level, 2014* (Luxembourg: EUR-OP, 2014).
37 Fabio Kon et al., 'A Panorama of the Israeli Software Startup Ecosystem,' *Orit and Yuklea, Harry, A Panorama of the Israeli Software Startup Ecosystem (March 1, 2014)* (2014).
38 Inbal Orpaz, 'Preserving the Madness' in Idf Intelligence,' *Haaretz*, September 26 2013.
39 John Stone, 'Politics, Technology and the Revolution in Military Affairs,' *Journal of Strategic Studies* 27, no. 3 (2004).
40 OECD Science, Technology and Industry Scoreboard 2015: , (Paris: OECD Publishing, 2015), http://dx.doi.org/10.1787/sti_scoreboard-2015-en.

total R&D expenditure as a share of GDP is foreign, and increased from 28% in 2007 to 47% in 2011.[41] It mostly consists of direct BERD investment and competitive funding awarded by European Research Programmes. The ratio of foreign investment indicates the degree of internationalisation of business R&D and the country's attractiveness to foreign investors.[42]

Conflict-laden Israel hosts R&D centres from most major IT multi-national corporations (MNCs).[43] In the 21st century, the latest R&D centres came about as a result of an MNC acquiring an Israeli company start-up in the software and IT security niches.

D. Cyber security industry exports

In 2014, Israeli companies held almost 10% of the global cyber security market, valued at $60 billion in 2013 by Gartner. Israeli companies exported IT security solutions (mostly software) worth $6 billion, double the $3 billion of exports in 2013.[44] According to the Israel National Cyber Bureau (INCB) estimates, Israeli cyber security exports reached $3.5 billion in 2015, about 5% of the global cyber security market valued now at $75 billion.[45] The dynamic innovation continues; Israeli society produced some 300 cyber security start-ups in 2015, up from 150 in 2012.

E. Formal national cyber policies

Government Resolution 3611 issued on August 7, 2011 – *Advancing the national capacity in cyberspace* – is the first Israeli national cyber strategy. It was the result of an external expert review, the 2010 *National Cyber Initiative*.[46] In order to promote the strategy which sought to 'make Israel a top-five global cyber power by 2015', an advisory body *Mat'e ha-Cyber ha-Leumi* (the Israel National Cyber Bureau INCB) was established in the Prime Minister's Office.[47]

A national contact point for cyber security incidents, the Israel National Cyber Event Readiness Team (CERT-IL), has operated since 2015.[48] In 2016, cyber protection of the civilian sector beyond critical infrastructure has yet to be developed in Israel. Accepting the recommendation of Isaac Ben-Israel's 2014 task force, the government resolved on February 15, 2015 to establish a new *Rashut Le'umit le-Haganat ha-Cyber* (National Cyber Security Authority, NCSA) to enhance cyber security in the civilian sector.[49] Before 2014, academic cyber research was

41 In the EU, *foreign* R&D expenditure as a share of GDP averages 10%. 'Gross Domestic Expenditure on R&D, by Type, 2013,' (OECD Publishing, 2015).

42 Richard R. Nelson, *National Innovation Systems a Comparative Analysis* (New York: Oxford University Press, 1993).

43 Uzi de Haan, 'The Israel Case of Science and Technology Based Entrepreneurship: An Exploration Cluster,' in *Science and Technology Based Regional Entrepreneurship* Global Experience in Policy and Program Development, ed. Sarfraz A. Mian (Cheltenham, UK: Edward Elgar Publishing, Inc., 2011).

44 'Cyber-Boom or Cyber-Bubble? Internet Security Has Become a Bigger Export Earner Than Arms.'

45 Author's interview with government officials, 02/2016. The nominal decrease is explained by foreign (mostly American) firms acquiring Israeli exporters, for a total of $1.3 billion in 2015, almost double $700 in 2014.

46 Lior Tabansky and Isaac Ben Israel, 'The National Cyber-Strategy of Israel and the Incb,' in *Cybersecurity in Israel, Springerbriefs in Cybersecurity* (Springer International Publishing, 2015).

47 Government of Israel, 'Government Decision 3611: Promoting National Capacity in Cyber Space,' (Jerusalem, Israel: PMO Secretariat, 2011).

48 https://cert.gov.il/

49 Israel Prime Minister's Office, 'Cabinet Approves Establishment of National Cyber Authority ' http://www.pmo.gov.il/English/MediaCenter/Spokesman/Pages/spokeCyber150215.aspx.

dispersed and uncoordinated in Israel. The story of cybersecurity policy in Israel shows Israel gained cyber power without many formal elements.

The various IDF branches and units have been operating relevant technology and capabilities towards highly ambitious goals without a joint cyber command. The decision to establish one was announced in June 2015.[50]

F. Defence experience

Qualitative superiority is imperative in Israel's strategy.[51] Israel's extensive defence R&D stems from the strategy. Defence R&D probably contributes an additional 1.5% of the GDP.[52] The Israeli Air Force (IAF), C4I Corps, and Intelligence Corps have long embraced cyber technology to perform their missions. Brigadier General (Ret.) Pinchas Buchris, the then Director General of the Israeli Ministry of Defence, said in a 2007 interview:

> 'I can only say we're following the network attack technology with great care. I doubted this technology five years ago. But we did it. Now everything has changed. Any such capabilities are top secret'.[53]

5. STRATEGIC CYBER POWER: THE ENDS-WAYS-MEANS NEXUS REVEALED

The brief discussion on strategic thought, the Israeli grand strategy, the Israeli national innovation system performance, and defence experience laid out the foundation for the strategic analysis. But technological prowess alone does not create power, nor can it compensate for strategic mistakes. Cyber power can only be meaningful in context; when one applies the means towards one's goals and tests both in action.

A. Soft power: R&D, innovation, business and diplomacy

Education, science, and research are the enduring cornerstones of the Israeli strategy to gain a qualitative edge. The National Innovation System drives scientific and economic development as well as cyber defence capability. The government explicitly leverages the academic, business and defence sectors for soft power.[54] The research universities serve (albeit not on purpose) the strategic goal of achieving and maintaining a qualitative edge by consistently developing human capital and advancing fundamental scientific research and applied technology. The business sector serves (again, not on purpose) strategic goals beyond the evident economic sphere. Israel has been consistently using its technological advances for diplomatic purposes, its assistance to Africa and Asia since the 1950s being the prominent example.[55] Nowadays, PM Netanyahu offers Israel's technological and operational expertise to other countries to

50 Gabi Siboni and Meir Elran, 'Establishing an IDF Cyber Command,' INSS, http://www.inss.org.il/index.aspx?id=4538&articleid=10007.

51 Jacob Amidror, 'Israel's Strategy for Combating Palestinian Terror,' *JFQ: Joint Force Quarterly*, no. 32 (2002); Eilam, *Eilam's Arc: How Israel Became a Military Technology Powerhouse*; Shlaim, *The Iron Wall: Israel and the Arab World*; Tal, 'National Security the Israeli Experience'.

52 Tabansky and Ben-Israel, *Cybersecurity in Israel*.

53 David A. Fulghum, Robert Wall, and Amy Butler, 'Israel Shows Electronic Prowess,' *Aviation Week & Space Technology* 168(2007).

54 Author's interview with senior Ministry of Foreign Affairs and the Prime Minister's Office officials.

55 Michael Curtis and Susan Aurelia Gitelson, *Israel in the Third World* (New Brunswick, N.J.: Transaction Books, 1976).

counter the forces that exploit cyberspace to wage war against Western values. Academic and business performance also attracts foreign direct investment (FDI) in Israeli science and high technology.

Strategic analysis shows how universities and business develop soft power, applied for the strategic goals:

- Reduce the cyber threat;
- Develop a prosperous economy;
- Increase cooperation with like-minded nations;
- Gain diplomatic benefit.

B. Hard power: Stuxnet

Operation Olympic Games, which has been attributed to the USA and Israel, demonstrated the real-world feasibility of striking high value, heavily defended targets with bits alone.[56] Probably implanted in late 2007, *Stuxnet* malware was specifically written to infiltrate air-gapped[57] networks and silently disrupt industrial control systems (ICS).[58] *Stuxnet* slowly and stealthily damaged the nuclear enrichment process at the Natanz facility in Iran by reprogramming the Siemens-made programmable logic controller (PLC) to spin the motor out of the safe range.[59] Stuxnet was a precision-guided weapon; the payload was only executed when the target met all predetermined conditions.[60]

Stuxnet targeted the Iranian *means*, towards the top Israeli strategic goals:

- Reduce and postpone the nuclear threat by rendering useless at least 1,000 of the 9,000 IR-1 centrifuges deployed at Natanz in late 2009 and early 2010, and having the unexpected failure rate introduce profound insecurity throughout the Iranian nuclear project;[61] and
- Reduce cyber risks, as developing cutting-edge capabilities in the ICS realm can improve critical infrastructure protection.

The effectiveness of Stuxnet remains a source of heated scholarly and policy debates. Critics argue that Operation Olympic Games failed to stop Iran's nuclear weapons programme; others argue it increased Iran's determination to pursue it.[62] There is, however, substantial strategic logic in this use of cyber capability as an instrument of power. The 'end' was to harm capacity,

56 David E. Sanger, *Confront and Conceal: Obama's Secret Wars and Surprising Use of American Power* (New York: Crown, 2012).

57 In IT-security, air-gapped refers to a network secured to the maximum by keeping it (often physically) disconnected from other local networks and the Internet.

58 Kim Zetter, *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon* (New York: Crown, 2014).

59 Ralph Langner, 'Stuxnet: Dissecting a Cyberwarfare Weapon,' *Security & Privacy, IEEE* 9, no. 3 (2011); Zetter, *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon*.

60 Thomas Rid, *Cyber War Will Not Take Place* (London: Hurst, 2013).

61 David Albright, Paul Brannan, and Christina Walrond, 'Did Stuxnet Take out 1,000 Centrifuges at the Natanz Enrichment Plant?,' (Washington, DC: Institute for Science and International Security, 2010).

62 Ivanka Barzashka, 'Are Cyber-Weapons Effective?,' *The RUSI Journal* 158, no. 2 (2013); Randall R. Dipert, 'Other-Than-Internet (Oti) Cyberwarfare: Challenges for Ethics, Law, and Policy,' *Journal of Military Ethics* 12, no. 1 (2013); James P. Farwell and Rafal Rohozinski, 'Stuxnet and the Future of Cyber War,' *Survival* 53, no. 1 (2011); Jon R. Lindsay, 'Stuxnet and the Limits of Cyber Warfare,' *Security Studies* 22, no. 3 (2013).

not intent. Expecting to alter the strategic course of a 78-million-strong nation is surely beyond realism. The 'way', a novel clandestine sabotage operation, was aligned with the 'ends' and echoes Israel's strategic culture.[63] The 'means', a first-of-its-kind, destructive, stealthy, precision-guided cyber weapon, expressed Israel's longstanding focus on its qualitative edge. The attempted physically destructive, precision-guided, prolonged, stealthy cyber attack to delay the main strategic threat to Israeli national security fits the definition of cyber power.

C. On formal policies

Strategy requires that decision-makers formulate and clearly communicate long-term ends in a reiterative fashion. Mundane democratic politics – structural checks and balances; coalition politics; electoral cycles; public opinion campaigns and more – make it difficult, yet some national leaderships have recently performed this task. However, the findings of the Israeli case study suggest cautious optimism; the formal process is not *sine qua non*.

6. SUMMARY, THEORETICAL IMPLICATIONS AND FUTURE RESEARCH

This essay outlined a new interdisciplinary analytical framework that integrates strategic studies and innovation system studies for strategic analysis of cyber power, and applied it to the Israeli case study. Western cyber insecurity stems largely from the skewed focus on means. Strategy – seeking how *means* and *ways* serve the exercise of soft and hard power for the national *ends* – is the missing ingredient in cyber power. A democracy seeking cyber power should optimally engage in an iterative strategic process loop:

- Reassess its particular strategy to clarify desired *ends*;
- Design cyber *means* by which *ways* can feasibly serve the defined strategic *ends*, focusing on non-military aspects, innovation and soft power;
- Experiment with and implement cyber means; and
- Reassess and continue to seek improvement.

A. Summary of the findings

Israel effectively develops cyber technology in the National Innovation System. But mere possession of technology does not neatly translate into power. The true manifestation of cyber power is in the application of means to achieve political ends. Crucially, the new analytical framework allows an improved understanding of cyber power. Israel exercises cyber technology for soft and hard power to meet national ends:

- Reduce the cyber threats and risks through security efforts;
- Develop a prosperous national economy;
- Increase cooperation with like-minded nations;
- Gain diplomatic benefit; and
- Reduce the Iranian nuclear threat.

63 Strategic culture refers to a set of national beliefs, attitudes and norms towards the use of force. See Dima Adamsky, *The Culture of Military Innovation: The Impact of Cultural Factors on the Revolution in Military Affairs in Russia, the U.S., and Israel* (Stanford, Calif.: Stanford University Press, 2010) for a discussion of Israel's strategic culture.

Although strategy is best organised formally, the Israeli case shows a serendipitous process prior to 2011. Israel gained cyber power without many formal elements: an official national cyber strategy, a committed government agency to coordinate cyber activity, a unified military command, a national CERT, or a dedicated academic thrust. The enduring strategic necessity to maintain a qualitative edge in order to develop a safe and prosperous State of Israel is what drives innovation throughout academia, industry, and defence operations. Much of the innovation is now expressed throughout cyber technology, and used in soft power efforts seeking strategic goals. Opting for cyber hard power to delay the main strategic threat to Israeli national security is a definite manifestation of cyber power.

B. Theoretical implications

This analysis aims to advance scholarly efforts, rather than grade policy. It may be tempting to present some of the findings on Israeli technological prowess as lessons[64] and offer sensible recommendations; to promote innovation, to invest more in R&D, or attract foreign business investment, all while cutting corners to escape the bureaucratic quagmire. Such 'lessons' would repeat the major flaw in countless cyber power debates; the focus on *means* and *ways*.

Western cyber insecurity stems largely from a common pitfall: the skewed focus on means. A society's capacity for innovation is one of the central enablers of successful adaptation to change; it drives ways and means. The economics scholarship on NIS can also contribute to scholarly and policy cyber power efforts alike. Even when properly integrated into systems and implemented by trained personnel, cyber technology cannot erase the difficulties that impede strategic excellence.[65] Only when applied towards clear ends, can ways and means be assessed. Strategic thought shows that the focus on means will take a heavy toll on cyber power scholarship and policy alike. Developing and adhering to strategic ends-ways-means logic will facilitate a transition from cyber technology to cyber power.

C. Future research directions: venture beyond technology

Advancing cyber power requires venturing beyond means, beyond the core technical disciplines and defence circles. Fields as disparate as international relations, change management in organisations, public policy, psychology, and many others can contribute potentially crucial knowledge. I partially illustrated the value of two disciplines: strategic studies and the economics of innovation. Cyber presents special obstacles. Distinct separate professional and scholarly communities, which interact only intermittently, is the academic reality. Secrecy concerns also further inhibit cyber research. In this era of change, leaders and strategists cannot afford the scientists' luxury of seeing experiments through; they must act under uncertainty. An improved understanding of cyber power demands further cross-disciplinary research and policy efforts to integrate more elements into the analytical framework.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Blavatnik Interdisciplinary Cyber Research Centre (ICRC) at Tel Aviv University (TAU).

64 William C. Fuller Jr., 'What Is a Military Lesson?,' *Strategic Studies: A Reader* (2008).
65 Colin S. Gray, *Strategy and History* (2006).

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

The Cyber-Doom Effect: The Impact of Fear Appeals in the US Cyber Security Debate

Sean T. Lawson
Department of Communication
University of Utah
Salt Lake City, Utah, USA
sean.lawson@utah.edu

Haoran Yu
Department of Communication
University of Utah
Salt Lake City, Utah, USA

Sara K. Yeo
Department of Communication
University of Utah
Salt Lake City, Utah, USA
sara.yeo@utah.edu

Ethan Greene
Department of Communication
University of Utah
Salt Lake City, Utah, USA

Abstract: In the US cyber security debate, observers have noted there is a tendency for policymakers, military leaders, and media, among others, to use frightening 'cyber-doom scenarios' when making the case for action on cyber security (Dunn Cavelty, 2008, p. 2). Some have conjectured that these fictional cyber-doom scenarios, which exemplify 'fear appeals' in communication research, have the potential to undermine productive debate aimed at addressing genuine cyber security challenges (Valeriano and Ryan, 2015, pp. 7, 196). Yet, few have directly investigated the impacts of such rhetoric. Here, we assess the impact of cyber-doom scenarios by contextualising them within existing scholarship on the impact of fear appeals, a well-studied phenomenon in communication science. First, we review qualitative and quantitative research on fear appeals and their effects. Next, we report results of an empirical study that takes advantage of a nationally televised docudrama depicting a hypothetical cyber-doom scenario. Through content analysis of real-time responses to this docudrama on the social media platform Twitter, we assess the effects of this particular cyber-doom scenario on a large audience. Our findings suggest that the use of such extreme fear appeals in the absence of clearly communicated and efficacious information about how to respond to the threat is counterproductive, as they can lead to a sense of fatalism and demotivation to act. Thus, concerns that the use of cyber-doom scenarios could impair efforts to motivate appropriate policy responses to genuine cyber security threats are warranted.

Keywords: *cyber security, fear appeals, policy, discourse, communication*

1. INTRODUCTION

Concern about cyber security in the United States is growing, and there are ongoing public policy debates about how to respond to these challenges. How the United States chooses to respond will have profound impact on the future of the Internet, global civil rights, and international security. In this debate, observers have noted that there is a tendency for policymakers, expert commentators, and news media, among others, to use frightening 'cyber-doom scenarios' when making the case for action on cyber security (Dunn Cavelty, 2008, p. 2). These scenarios involve fictional tales of cyber attack resulting in mass destruction or even total economic and social collapse. Although they are not necessarily reflective of actual cyber threats facing the nation (Clapper, 2015), such scenarios are still common in US public policy discourse. Some have conjectured that these fictional cyber-doom scenarios, which are an example of 'fear appeals' in communication research, have the potential to undermine productive debate aimed at addressing genuine cyber security challenges (Valeriano & Maness, 2015, pp. 7, 196). Yet, few have directly investigated the impacts of such rhetoric.

Our paper assesses the impacts of cyber-doom scenarios by contextualising them within existing research findings on the impacts of fear appeals, a well-studied phenomenon in communication science. The goal of this paper is to provide an assessment of the degree to which cyber-doom scenarios could impair efforts to motivate appropriate policy responses to genuine cyber security threats. Assessments like those provided in 2015 by Director of National Intelligence (DNI) James R. Clapper indicate that the frightening rhetoric of cyber-doom scenarios is likely not an accurate reflection of real cyber threats. But can such rhetoric be dangerous in its own right?

In this essay, we first review qualitative and quantitative communication research on fear appeals and their effects. Then, we supplement our review by presenting preliminary results of an empirical study that takes advantage of a nationally televised docudrama depicting a hypothetical cyber-doom scenario. The fictional National Geographic Channel docudrama, *American Blackout*, aired in October 2013 and reached roughly 86 million American households (Seidman, 2015). Through content analysis of real-time responses to this docudrama on the social media platform Twitter, we assess the effects of this particular cyber-doom scenario on a large audience. Our findings suggest that the use of such extreme fear appeals in the absence of clearly communicated, obtainable, and efficacious information that viewers can use to help address the problem are counterproductive as they can lead to a sense of fatalism and de-motivation to act. Thus, concerns that the use of cyber-doom scenarios could impair efforts to motivate appropriate policy responses to genuine cyber security threats are warranted.

2. CYBER-DOOM SCENARIOS

For more than a decade, scholars have noted the use of cyber-doom scenarios by news media, expert commentators, and policymakers when talking about cyber security (Debrix, 2001; Weimann, 2005; 2008; Stohl, 2007; Conway, 2008; Dunn Cavelty, 2008, p. 2; Lawson, 2013a;

Valeriano & Maness, 2015). Perhaps the most influential recent use of a cyber-doom scenario by a policymaker occurred in 2012 when former US Secretary of Defense Leon Panetta warned about the possibility of what he termed 'cyber Pearl Harbor' in which coordinated cyber attacks wreak unprecedented destruction and chaos on the nation (Panetta, 2012). But Secretary Panetta was not the first or the last to contemplate such scenarios. In fact, the Pearl Harbor analogy dates back to 1991 when computer security expert and novelist Winn Schwartau warned about the threat of an 'electronic Pearl Harbor' (Schwartau, 1991). In the intervening years, analogies, metaphors, and scenarios positing cyber attacks with effects akin to military attacks, natural disasters, and nuclear weapons, have been common in the cyber security debate (Debrix, 2001; Conway, 2008; Clarke & Knake, 2010; Lawson, 2013a). In 1994, the influential futurist and theorist of the Information Age, Alvin Toffler, warned that terrorists could cyber attack the World Trade Centre and crash the US economy (Elias, 1994). In 1999, Fox News ran a documentary, *Dangers on the Internet Highway: Cyberterror*, warning of the possibility of catastrophic cyber attacks (Debrix, 2001; Conway, 2008). Eleven years later, CNN ran a televised war game called *Cyber.Shockwave*, which contemplated the implications of a massive cyber attack. That same year, Richard Clarke and Robert Knake began their book, *Cyber War*, with a tale of cyber attack crippling all US critical infrastructure and killing thousands in only a matter of minutes (Clarke & Knake, 2010). Others have speculated that cyber attacks could be as devastating as the 9/11 terrorist attacks (Martinez, 2012), the 2004 Indian Ocean tsunami (*The Atlantic*, 2010), Superstorm Sandy (Meyer, 2010), or the Fukushima nuclear disaster (Rothkopf, 2011). One former policymaker even warned that cyber attacks could pose a threat to all of global civilisation (Adhikari, 2009).

Cyber attacks against critical infrastructure are certainly not impossible, and we have seen examples of cyber attacks causing physical damage or destruction, the Stuxnet attack on Iranian nuclear facilities being perhaps the most prominent example thus far. Nonetheless, we have not seen attacks that come even close to causing the kinds of chaos and destruction contemplated in cyber-doom scenarios. Indeed, in the face of persistent warnings of cyber-doom, the US Director of National Intelligence told Congress twice in 2015 that such 'Cyber Armageddon' scenarios are not reflective of the real cyber threats facing the nation (Clapper, 2015). However, despite this clear rejection of cyber-doom scenarios by the nation's top intelligence official, warnings of a 'cyber Pearl Harbor' or 'cyber 9/11' persist among policymakers, commentators, and journalists. In February 2015, NSA Director, Admiral Michael Rogers, claimed that the hack of Sony Pictures the previous year constituted a 'cyber Pearl Harbor' of the kind Secretary Panetta had warned about in 2012 (Lyngaas, 2015). That same month, in a speech on cyber security, President Barack Obama urged listeners 'to imagine' cyber attacks that 'plunge cities into darkness' (Obama, 2015). In August 2015, Senator Susan Collins (R-ME) urged the passage of the Cybersecurity Information Sharing Act of 2015 'to reduce the likelihood of a cyber 9/11' (Collins, 2015). The legislation later passed (Pagliery, 2015). Finally, veteran journalist and television news personality Ted Koppel made headlines with his October 2015 book warning of the possibility of catastrophic cyber attacks on the power grid (Koppel, 2015). Indeed, since DNI Clapper's February 2015 statement to Congress, at least two dozen articles have appeared in major US newspapers warning of either 'cyber Pearl Harbor' or 'cyber 9/11'.[1] It is perhaps unsurprising, therefore, that cyber terrorism ranked second only to government

1 Based on a search of LexisNexis Academic Universe database of 'US Newspapers' on 24 February 2016. Inclusion of broadcast transcripts, wire services, and online news sources not covered by LexisNexis would certainly turn up even more instances.

corruption in a survey of average Americans' fears in 2015, even beating traditional terrorism (Ledbetter, 2015).

Critics of the persistent use of cyber-doom scenarios point to several potential dangers of relying too heavily on such worst case thinking. Framing cyber threats in extreme terms invites a militarised response that may be ineffective and even counterproductive for dealing with the cyber threats that we do face (Lewis, 2010). In the first case, it is not at all clear that the military is the appropriate institution for dealing effectively with the kind of broad cyber threat to private intellectual property and personal information identified by DNI Clapper and his predecessors (Lawson, 2013b). More concerning, however, is the possibility that the types of policies and responses that worst case thinking promotes are actually counterproductive. Militarised cyber security policies by the US could undermine its own policy of promoting Internet freedom around the world. In an interconnected environment such as cyberspace, the kinds of offensive actions often contemplated or, in some cases already undertaken, can 'blow back' onto the party who initiated those actions, leading to unintended, negative consequences (Dunn Cavelty, 2008, p. 143; Lawson, 2015; Dunn Cavelty & Van Der Vlugt, 2015). In other cases, such framing might encourage defensive actions that are ineffective or even counterproductive (Ball et al., 2013; Gallagher & Greenwald, 2014; Schneier, 2014). There also exists the possibility that worst case, cyber-doom thinking could distract from and lead to a sense of complacency about the more mundane, but realistic, cyber threats that we do face (Debrix, 2001, p. 156; Lewis, 2010, p. 4; Lawson, 2012). Finally, some worry that worst-case, cyber-doom thinking and the militarised responses it promotes could end up as a self-fulfilling prophecy, leading to conflict escalation where non-physical cyber attacks escalate to physical warfare, or even to the kinds of preventive war scenarios witnessed in the 2003 US invasion of Iraq (Furedi, 2009; Thierer, 2013; Blunden & Cheung, 2014; Valeriano & Maness, 2015).

3. FEAR APPEALS

Cyber-doom scenarios are an example of the use of fear appeals to raise awareness of, and motivate a response to, cyber security problems. In general, scholarship indicates that while fear appeals can be effective and ethical forms of argument, they are prone to failure, to producing counterproductive effects, and to being used in ways that are detrimental to political deliberation in a democracy.

A fear appeal is a kind of argument that attempts to persuade or influence through the use of 'warning[s] that some bad or scary outcome will occur if the respondent does not carry out a recommended action' (Walton, 2000, p. 1). Fear appeals can take a number of forms. Cyber-doom scenarios, however, most closely resemble the form of fear appeal that works based on invoking uncertainty about a possible future. In this form, '[s]ome dangerous event that, it is said, might happen in the future, raises gloomy foreboding and fears related to the uncontrollability of what could possibly happen in an uncertain world. Fear appeal arguments [of this type] trade on uncertainty about a possible future sequence of events that might be set into motion once a step in a certain direction is taken' (Walton, 2000, pp. 14-15) or, we might add, *not taken*.

There is a long tradition of studying such arguments. Rhetoricians, logicians, and scholars of argumentation have been concerned with the effectiveness of such arguments, but also their logical structure, variations, and ethics. These scholars have traditionally argued that fear appeals are fallacious, unethical, or both because they rely on appeals to emotion or, in some cases, overt threats such as in a classic protection racket. However, more recent scholarship has questioned the notion that fear appeals are always fallacious or unethical (Walton, 2000; Pfau, 2007). For example, Pfau (2007) examines Aristotle's advice about how to effectively employ appeals to fear. Aristotle advised that to be effective one must convince the audience that the threat is painful and destructive, that it is near, that it is contingent (preventable or controllable), and buoy the courage of the audience to act. Pfau (2007, pp. 231-233) argues that, for Aristotle, fear appeals can be effective and ethical if they are employed to call attention to a real danger, serve to open deliberation, and encourage appropriate responses, as opposed to closing discussion and coercing a pre-determined response. He also notes that Aristotle warned against the use of 'overpowering fears', which could inspire 'flight or resignation and inaction' instead of appropriate responses (Pfau, 2007, p. 227).

Social scientists have posited models of fear appeals that bear a close resemblance to the one offered by Aristotle. In the model proposed by Rogers (1975), there are three components of fear appeals: (i) the severity of the threat; (ii) the probability of occurrence of the threat; and (iii) the efficacy of a response (see also Maddux & Rogers, 1983). More recently, models of fear appeal messages have been said to be composed of four components, two related to the threat and two related to the recommended response (Witte, 1994). The threat components convey the ideas that the threat is particularly harmful (severity) and that the listener is at risk of experiencing these harmful effects (susceptibility). The response components convey the ideas that the recommended response will be effective (response efficacy) and that the listener is capable of carrying out the response (self-efficacy; Witte, 1994, p. 114). The study of why fear appeals succeed or fail has been prominent in health communication, particularly among those concerned with how to promote healthy behaviours and discourage unhealthy behaviours. More recently, researchers in the field of information security have looked to health-related fear appeals research to guide their own work on promoting better security practices among computer users (Boss, et al., 2015).

In general, studies of the effectiveness of fear appeals in health communication and information security have largely confirmed Aristotle's advice; some use of fear is helpful, but too much is counterproductive. Success occurs when listeners engage in danger control behaviours, those that reduce the threat. Failure occurs when listeners engage in fear control behaviours, those that reduce their feelings of fear but do nothing to prevent, or sometimes even increase the risk of, the threat. Initially, researchers hypothesised that the greater the fear elicited in the fear appeal message, the greater the likelihood of message success. That, however, turned out not to be the case. Instead, researchers have found that fear only works up to a certain point. Too much fear can actually be counterproductive. Aristotle's 'contingency' and 'courage' seem to be key to message success. Listeners cannot only be scared into action. They must also believe that something effective can be done to address the threat and that they are capable of carrying out the necessary response. That is, threat components of the fear appeal message must be accompanied by, and in balance with, convincing response components for the message

to succeed (Witte, 1994; Peters, et al., 2013). Researchers in information security have only recently begun to explore the role of fear appeals in promoting better security practices, but this early work tends to agree with the findings from health communication (Doohwang, et al., 2006; Herath & Rao, 2009; Pfleeger & Caputo, 2011; Siponen, et al., 2014; Boss, et al., 2015).

In addition to the effectiveness of fear appeal messages, scholars of rhetoric, logic, and argumentation have also explored the ethical and normative aspects of fear appeals. This work lends support to the concerns raised about possible negative effects of cyber-doom scenarios. Although recent scholarship rejects the traditional idea that fear appeals are always fallacies and are unethical, this work still maintains that fear appeals can be dangerous. For example, Walton (2000, p. 199) argues that these arguments can serve as 'a potent obstacle to free democratic political deliberations and open critical discussions of political issues'. He describes various cases in which fear appeals are weak, unethical, or even fallacious forms of argument that are 'destructive to the democratic process'. These cases include instances where speakers resort to fear appeals because of weak evidence or weak ties between their premises and conclusions. That is, they use fear or threat as a shortcut to prematurely close down deliberation and get their way (Walton, 2000, pp. 188-191). Similarly, fear appeals can be fallacious and unethical when they rely on deception. In these cases, the speaker knows that the fear or threat is not supported by, or that it is even contradicted by, the evidence (Walton, 2000, pp. 193-194). Finally, fear appeals can also be unethical and perhaps fallacious when they are used as a tool of misdirection or distraction in political deliberation, taking attention away from other, relevant issues or considerations and focusing attention instead on one, emotionally charged issue (Walton, 2000, p. 200).

4. AMERICAN BLACKOUT

The fictional *National Geographic Channel* docudrama, *American Blackout*, aired in October 2013 and reached roughly 86 million American households (Seidman, 2015). In addition to depicting a cyber-doom scenario in detail, this programme is exemplary of the blurring of distinctions between news and entertainment media that some have argued are central to the emergence of a culture and politics of fear in the last several decades (Altheide, 2002, 2006; Glassner, 1999). Thus, this programme and the responses that it elicited on social media are valuable for understanding how traditional and new media contribute to the articulation of cyber security-related fears and audience responses to the communication of those fears.

We collected the responses to the show on the social media platform, *Twitter*. Tweets with the hashtag #AmericanBlackout were collected on the first night that the show aired and for about 12 hours afterwards using a free tool called *Twitter* Archiving Google Spreadsheet (TAGS) v.5. This tool uses the *Twitter* API to collect into a spreadsheet tweets meeting a certain search criteria.[2] Though the program reached 86 million U.S. homes, gauging viewer responses to the program using more traditional methods would require knowing which of the 86 million homes, and who in them, actually viewed the program so that a survey could be conducted. However, using Twitter responses that included the hashtag #AmericanBlackout had the advantage of providing a more direct route to a group of people who presumably watched the program or

2 For more information about his tool, see https://tags.hawksey.info/ (accessed December 29, 2015).

were aware of it. This collection method resulted in 16,501 tweets. We content analysed a random sub-sample (10 percent) of the collected tweets for preliminary analysis. Of the 1,650 tweets in the sub-sample, one tweet was not in English and was thus excluded from analysis. In accordance with models of fear appeals, we content analysed the tweets for susceptibility, severity, and efficacy of responses. Because we were interested in the type of responses to cyber-doom scenarios, we examined the tweets for preventative and reactive responses, that is, tweets mentioning responses to prevent or react to a cyber-doom scenario like the one depicted in the show. It is important to note that these data represent viewers' responses to the docudrama and are thus people's *perceptions* of the threat and recommended responses. In addition to perceptions of threat, efficacy, and types of recommended responses, we also coded any expressions of fatalistic reactions or avoidance in each tweet, such as tweets where individuals expressed the idea that nothing could be done or a desire to avoid thinking about such a scenario. Finally, as tweets can be original (created and posted by the user) or re-posted content, we felt it important to quantify how many tweets were re-tweets or modified tweets, which are re-posts that are altered in minor ways. Descriptions of the variables coded in this study and examples of tweets can be found in Table 1. Two independent coders each read and coded the full sample of 1,649 tweets. Disagreements between coders were reconciled through discussion. Of the 1,649 tweets coded in our preliminary analysis, 1,157 (70.2 percent) were re-tweets or modified tweets. Although the majority of tweets were not original, users are likely to re-post content as a way to engage with other users tweeting about the show and further share content they believe worthy of dissemination.

TABLE 1: DESCRIPTION OF CODED VARIABLES AND EXAMPLES FROM TWEETS CONTAINING #AMERICANBLACKOUT

Variable	Definition	Examples
Susceptibility	Expression that he/she is likely to be in such a scenario	When will the real #americanblackout happen?
Severity	Expression that the threat of cyber-doom is harmful and/or large	Im freaking out right now im worried about my kids :(#americanblackout
Presence of response	Tweet expresses that individual perceived some response to threat	#americanblackout is petrifying. I will now become a doomsday prepper.
Efficacy of response	Tweet expresses whether the perceived response will work	#americanblackout M.R.E I had five cases and gave them away wish I hadn't now. M.R.E are the way to go they last long for years
Self-efficacy	Belief about whether user is capable of carrying out responses and/or cope with the threat	#americanblackout after this show I am surely becoming a doomsday prepper / survivalist when the s*** hits the fan what are you going to do?
Preventative government response	Expression of government response that is preventative	#americanblackout is so freaking scary omg i would die like no. the government better ensure that NEVER happens or im movin to canada
Preventative personal response	Expression of personal response that is preventative	

Variable	Definition	Examples
Preventative other response	Expression of preventative response not associated with personal or government actions	Lets all cross our fingers and pray this never happens, lol I'll be looking like a cave women.. if I survive. :o #americanblackout
Reactive government response	Expression of government response that is reactive	FEMA repeats orders for millions of body bags
Reactive personal response	Expression of personal response that is reactive	I need to go buy a few gallons of water tomorrow. It might get real, soon. #AmericanBlackout
Reactive other response	Expression of reactive response not associated with personal or government actions	Learn how to get prepared rationally, by real people. Not Scared. #preppertalk chat Daily-6PM Eastern. #AmericanBlackout
Fatalistic reaction or avoidance	Expression of inability or unwillingness to act or respond to threat	Honestly, I don't think I would survive. #americanblackout
Re-tweets	Tweets that contain 'RT,' 'MT,' or quotations marks around content	RT @Aj_Dreww: This #americanblackout is freaking me out...

As a fear appeal, *American Blackout* begins with two epigraphs that establish the supposed severity, susceptibility, and credibility of the threat depicted in the programme. The first, a quote from Dr. Richard Andres of the US National War College, asserts, '[a] massive and well-coordinated cyber attack on the electric grid could devastate the economy and cause a large-scale loss of life'. The second explains, '[t]he following program draws upon previous events and expert opinions to imagine what might happen if a catastrophic cyber attack disabled the American power grid'. The implication is that such an attack is a possible scenario that would severely affect the entire nation.

Much of the remainder of the show repeatedly reinforces these themes of severity and susceptibility as we see cell phone videos taken by average people documenting their attempts to survive what turns into a ten day blackout. These personal cell phone videos are interspersed with news footage, which helps to provide the big picture view of the cyber attack's effects. In this scenario, no one is untouched by these effects, which in just three days includes violence, looting, rioting, and loss of life. Early in the programme, one citizen tells his camera that the United States has become 'a Third World country'. Later, another talks about society no longer existing, and even the President admits that government cannot 'keep society afloat.' The implication is clear: Electricity and society are one and the same; without the former, the latter quickly ceases.

Despite the show's attempt to portray such a scenario as frighteningly harmful and likely, our analysis of *Twitter* responses shows that only 35.8 percent of tweets contained expressions of perceived susceptibility to cyber attack and only 26.3 percent of tweets contained expressions of perceived severity of the threat of cyber attack. However, a smaller proportion contained expressions of both severity and susceptibility (21 percent), while 48.8 percent of tweets did not contain mentions of either dimension.

Though *American Blackout* quickly, clearly, and repeatedly establishes the threat components of its fear appeal message, the programme does not clearly articulate a recommended response for preventing the threat of cyber attack or for mitigating the effects should such an attack occur. In the show, any recommended responses that are even implied are focused on government, individual, family, or small group responses in the aftermath of the attack. There are no depictions of responses at a level between the government and the individual or small group. This is exemplified in the tweets about the show, only 20.1 percent of which contained mentions of recommended responses to a cyber attack.

Where government is concerned, the programme depicts centralised and militarised responses, such as military and riot police using force to quell riots and looting, as well as declaring a state of emergency in which the federal government takes centralised control of all food and water supplies. However, the majority of tweets did not mention government responses. Only 5.0 and 0.2 percent of tweets mentioned preventative (e.g., power grid drills) and reactive (e.g., declaring a state of emergency) government responses, respectively. In the programme, individuals, families, and small groups mitigating the effects of the attack for themselves are largely portrayed as helpless to prevent or effectively react to such a cyber-doom scenario. This is borne out in the content analysis where none of the 1,649 tweets coded contained expressions of preventative actions that individuals or small groups could take. In fact, almost all (99.1 percent) of the tweets contained no expression of self-efficacy at all.

In the show, there is one exception with regards to individuals' ability to mitigate the threat; a family of so-called 'preppers.' These are people who prepare for doomsday by planning to 'bug out' to a safe location stocked with food, water, and weapons. In the programme, the prepper family does just that, withdrawing from society to an undisclosed location in the mountains, refusing to help their neighbours, dressing in military garb, and drawing weapons on a neighbour who asks for some food and water. Throughout the show, an advertisement for the *National Geographic Channel* show, *Doomsday Preppers* appears often in the upper right corner of the screen, another tacit endorsement of 'prepping' as the only possible response to a cyber-doom scenario. Roughly 7.5 percent of tweets contained reactive responses to *American Blackout*, typically expressing intentions to become preppers ('I'm about to become a doomsday prepper after watching #americanblackout this is mad scary').

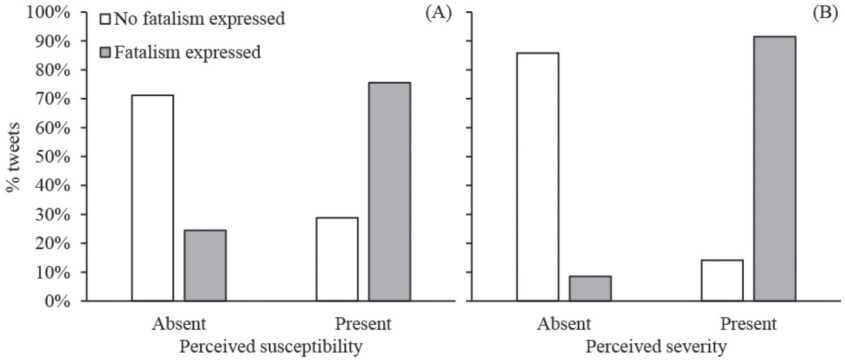
Even though the docudrama seems to tacitly endorse militarised and centralised government responses and 'prepping' on the part of individuals and families, neither of these responses is depicted as particularly effective. In the first instance, a week into the crisis the President must admit that government efforts are not sufficient and asks for assistance from the United Nations. In the second, even the prepper family barely survives. When the show comes to an end, they are in an armed standoff with a group trying to steal their supplies and are only saved when the power comes back on, which is represented by the sudden sound of a ringing iPhone. Just as suddenly, the standoff ends as society/electricity is restored. Equally as abruptly, the show comes to a close. As suddenly as lack of electricity destroyed society and led to disaster from which there was no escape or response, society was just as quickly restored and all was right with the world once again.

As a fear appeal, we can say that this cyber-doom scenario has strong and clear threat components that articulate a clear vision of the severity and susceptibility of a massive cyber attack. Moreover, this fear appeal is much less clear when it comes to offering recommended responses, which are arguably entirely absent. Likewise, the message does little to address either response or personal efficacy, largely depicting any implied responses as ineffective or, in the case of 'prepping', requiring years of work and a great amount of resource to achieve results that are uncertain at best in terms of their effectiveness.

It is perhaps unsurprising, therefore, that viewers were more likely to tweet about the threat components of the show than about the response components. Nonetheless, the volume of viewer mentions of severity or susceptibility was still quite low, which may be read as an indicator that they did not perceive the scenario depicted in the show to be believable. This is important because, as discussed below, perceptions of the believability of the threat components influence perceptions of the response components of the message. Certainly, a small number of viewers explicitly rejected the threat components outright. Similarly, when response components of the show were mentioned at all, they tended to mention government action to prevent the threat, or individual responses focused on reacting to the threat by 'prepping'. Viewers did not perceive that there was anything effective that they could do to prevent such a scenario.

As fear appeals are a persuasive messaging technique, the outcomes are typically message acceptance or rejection. Using the Extended Parallel Process Model (EPPM) of information processing, Witte (1994) posits that message acceptance is motivated by a need for protection against the threat, while rejection is driven by defensive motivation. Defensive motivation occurs 'when perceived threat is high and perceived efficacy is low, and produces message rejection responses such as defensive avoidance or reactance' (Witte, 1994, p. 116). Our content analysis found the majority of tweets (83.6 percent) contained no expressions of defensive motivation. The remaining 16.4 percent contained fatalistic or avoidance reactions characterised by an inability to deal with the threat. Examples include '#AmericanBlackout If this really happens I can't deal' and 'Uhmm I am ok with not watching #AmericanBlackout. Don't really want to imagine all the terrible things that would happen.' We used chi-square tests to examine the relationships between perceived susceptibility and expressions of fatalism, as well as perceived severity and expressions of fatalism. We found these relationships to be significant (perceived susceptibility-fatalism: $\chi^2 = 211.54$, df = 1, $p \leq .00$; perceived severity-fatalism: $\chi^2 = 675.17$, df = 1, $p \leq .00$). Among tweets that had no evidence of defensive avoidance (83.2 percent of all tweets), 71.2 percent contained no expression of susceptibility to the threat. Conversely, among tweets that expressed fatalism (16.8 percent of all tweets), 75.6 percent contained some expression of defensive motivation in the form of fatalism or avoidance of the scenario. A similar pattern is observed in the relationship between perceived severity and fatalism (Figure 1). In other words, our data suggest that perceived severity and susceptibility are positively related to expressions of defensive motivation.

FIGURE 1: CROSS-TABULATIONS OF PROPORTIONS OF EXPRESSIONS OF PERCEIVED SUSCEPTIBILITY (A) AND SEVERITY (B) WITH THOSE OF DEFENSIVE MOTIVATION. IDENTICAL COLORED BARS IN EACH PANEL TOTAL 100 PERCENT



Given the evidence, it is clear that the frightening rhetoric of cyber-doom scenarios, such as the one depicted in *American Blackout*, can be counterproductive to addressing real cyber attack threats, particularly if such messaging leads people to discount or downplay the potential threat.

5. CONCLUSION

The findings presented here lend support to concerns that the use of cyber-doom scenarios could have a negative impact on our ability to raise awareness of, and appropriately respond to, actual cyber security challenges. Existing scholarship in communication on the use of fear appeals has consistently demonstrated that such arguments can be counterproductive, unethical, or both, when they rely too much on raising fear of a threat, while simultaneously failing to offer the audience effective responses, or at least promote deliberation about possible responses. These findings are borne out in our preliminary assessment of instantaneous viewer responses on social media to the cyber-doom scenario depicted in *American Blackout*. Viewers were more likely to respond to the threat components of the message than to the response components, which makes sense given the strength of the threat depicted in *American Blackout* and weakness of the efficacy component. Nonetheless, the volume of responses to the threat component was still low, a potential indicator that most viewers did not find the scenario believable. Viewer responses that did mention the threat components were more likely to also express a sense of fatalism about the threat. Likewise, few responses indicated that viewers believed that there was something efficacious that either they or the government could do to prevent or respond to the scenario depicted in *American Blackout*.

Despite our preliminary analysis, it is difficult to determine whether *American Blackout* was successful as a fear appeal message. Judging the success of the message depends on knowing its intended effects. However, beyond the obvious business goals of any media organisation (viewership and advertising revenue), the goals of this program remain uncertain. However,

the most common goals for fear appeals messages in general are the promotion of particular preventive or reactive responses to, or the raising of awareness about, a threat. In the first case, if the goal of this particular fear appeal was to promote a particular preventive or reactive response, it seems to have failed. The vast majority of viewer responses did not express a perceived recommended response. In the second case, if the goal of the program was merely to raise awareness, to call attention to the problem rather than promote any specific response, there are still at least two possible problems. First, scenarios like the one in the program could raise awareness of the wrong problem; second, if audiences find a frightening, over-the-top depiction of cyber threats to be unbelievable, they may be more likely to discount or downplay all messages about cyber threats, even ones that are more realistic. There is another possible intended effect, however; the possible intent of using such scenarios is to scare audiences into passively acquiescing to government actions to combat cyber threats. If this was the intent behind *American Blackout*, then the fatalism exhibited by those who responded to the threat component of the message may indicate that this fear appeal was successful after all. If this were the case, then this message would meet the definition of one that is fallacious, unethical and thus deleterious to deliberation and decision-making in a democracy.

As we have noted throughout, this work represents a preliminary assessment of an extreme cyber-doom scenario and audience responses to it. More work is needed to analyse a larger sample of responses, as well as supplementary media produced as part of the *American Blackout* programme. These might include accompanying website and articles, expert interviews, infographics, and *National Geographic Channel* social media messaging, and responses to the programme in other venues such as blogs or news stories. Similarly, more work is needed to assess whether these findings hold when cyber-doom scenarios are depicted in different media and by different sources, such as government officials. Finally, not all cyber-doom rhetoric involves explicit depictions of worst-case scenarios like the one in *American Blackout* or Secretary Panetta's 2012 'cyber Pearl Harbor' speech. Indeed, cyber-doom rhetoric often involves the more subtle use of analogies and metaphors that imply the possibility of cyber attacks approximating the effects of military attacks or natural disasters but do not provide explicit depictions of those effects. More work is needed that seeks to measure empirically the effects of this more subtle form of cyber-doom rhetoric.

At minimum, however, the findings of this study lend support to concerns about the possible negative effects of cyber-doom rhetoric and should thus encourage policy makers, commentators, industry experts, journalists, and other cyber security advocates to be more cautious in their messaging strategies. Indeed, our study provides insights into recent findings from the Chapman University Survey of American Fears 2015, which found that fear of 'cyber-terrorism' ranked second only to government corruption in a list of top ten fears that average Americans said make them 'afraid' or 'very afraid' (Ledbetter, 2015). We noted above that one danger of cyber-doom rhetoric is that it can raise awareness of the wrong problems. The Chapman Survey may provide evidence that this is indeed occurring on a wider scale. For example, the survey showed that 'cyber-terrorism' (an as-yet hypothetical concern) ranked higher than other concerns directly related to actual cyber threats. These included tracking of personal information by corporations and government, identity theft, 'running out of money', and credit card fraud, all of which are related to the actual, low-level cyber threats over time to personal, corporate, and government

data that DNI Clapper and so many others have consistently identified as representing the true cyber threat. Indeed, cyber-terrorism outranked traditional terrorism even at a time when ISIS was on the march in the Middle East and North Africa.

A second possible danger of fear appeals in general, and cyber-doom rhetoric in particular, identified in the literature and in our study, was a tendency towards fatalism and demotivation when threats are overemphasised relative to effective responses. It is potentially significant to note, therefore, that although cyber-doom rhetoric has been prominent in US public policy discourse about cyber security, and 'cyber-terrorism' was ranked second among American's top fears in 2015, we have seen very little government action on cyber security even as experts continue to downplay the threat of cyber-doom. For example, in February 2016, former Director of the National Security Agency and the Central Intelligence Agency, General Michael Hayden, echoed DNI Clapper's 2015 assessment of the cyber threat when he told the *Wall Street Journal* that fear of 'cyber Pearl Harbor, digital 9/11, catastrophic attack' are misplaced, and that the only piece of cyber security legislation passed thus far – the Cybersecurity Information Sharing Act of 2015 – is essentially too little, too late, and leaves businesses and individuals largely responsible for their own cyber defence (Bussey, 2016). The combination of what we know from the existing fear appeals literature, the findings of our study, and the results of the Chapman survey indicate that the persistence of cyber-doom rhetoric may help to explain this lack of substantive progress in addressing the most widespread cyber threats that Americans actually face.

This suggests lessons for policymakers, experts, news media, and others responsible for crafting and communicating responses to cyber threats. These actors should think much more consciously and carefully about the intended effects of messages meant to communicate the cyber threat to their peers and the wider public.

In turn, such messages should be more carefully crafted and targeted. There has been a tendency in cyber-doom rhetoric to conflate very different threats into one monolithic and more frightening cyber threat in an attempt to raise awareness and motivate a response (Lawson, 2013b). However, there is not just one threat, but many, each of which may need to be addressed by different actors, including businesses and even average computer users (Singer, 2016). As Peter Singer recently noted, basic 'cyber hygiene' 'would stop 90 percent of cyber attacks, and help to keep all of us safe' (Singer, 2016). Indeed, he is not the first to have suggested that a public health approach to cyber security is necessary (Charney, 2010).

As Gen. Hayden notes, in the absence of sufficient government action on cyber security, organisations and individuals must do more to defend themselves. As Singer and others note, this may actually be the best approach. In either case, communicating specific cyber threats in a way that encourages organisations and individuals to take action, not merely to wait for government, will be crucial to promoting improved cybersecurity. Doing that will require policymakers, experts, and news media to tone down the cyber-doom rhetoric and instead communicate clear and specific messages about specific, realistic threats and, most importantly, what audiences of such messages can do themselves to help address those threats. From Aristotle to recent social science, we know that more fear is not better when trying to motivate audiences to action. This

applies to cyber security as much as to wearing one's seat belt or quitting smoking. In short, those responsible for effectively communicating cyber threats would do well to heed a version of Michael Pollan's dictum for healthy eating attuned to communicating threats: 'Use fear. Not too much. Focus on effective and obtainable responses'.[3]

REFERENCES

Adhikari, R. (2009, 2 December). Civilization's high stakes cyber-struggle: Q&A with Gen. Wesley Clark (ret.). *TechNewsWorld*. Retrieved from http://www.technewsworld.com/story/Civilizations-High-Stakes-Cyber-Struggle-QA-With-Gen-Wesley-Clark-ret-68787.html?wlc=1259861126&wlc=1259938168&wlc=1290975140.

Altheide, D. L. (2002). *Creating fear: news and the construction of crisis*. New York: Aldine de Gruyter.

Altheide, D. L. (2006). *Terrorism and the politics of fear.* Lanham, MD: AltaMira Press.

Ball, J., Borger, Julian, & Greenwald, G. (2013, 06 September). Revealed: How US and UK spy agencies defeat internet privacy and security. *The Guardian*. Retrieved from http://www.theguardian.com/world/2013/sep/05/nsa-gchq-encryption-codes-security .

Blunden, W., & Cheung, V. (2014). *Behold a pale farce: cyberwar, threat inflation, & the malware-industrial complex*. Waterville, OR: Trine Day.

Boss, S. R., Galletta, D. F., Lowry, P. B., Moody, G. D., & Polak, P. (2015, forthcoming). What do systems users have to fear? Using fear appeals to engender threats and fear that motivate protective security behaviors. *MIS Quarterly*.

Bussey, J. (2016, 9 February). Gen. Michael Hayden gives an update on the cyberwar. *Wall Street Journal*. Retrieved from http://www.wsj.com/articles/gen-michael-hayden-gives-an-update-on-the-cyberwar-1455076153.

Charney, S. (2010). *Collective defense: Applying public health models to the Internet*. Redmond, WA: Microsoft Corp.

Clapper, James R. (2015, September 10). "Statement for the Record: Worldwide Cyber Threats." House Permanent Select Committee on Intelligence. Retrieved September 10, 2015 from http://www.dni.gov/files/documents/HPSCI%2010%20Sept%20Cyber%20Hearing%20SFR.pdf.

Clarke, R. A., & Knake, R. (2010). *Cyber War: The Next Threat to National Security and What to Do About It*. New York: HarperCollins.

Collins, S. (2015, 06 August). Senator Collins continues to sound the alarm on the urgent need to bolster cybersecurity. *Press Release, Office of Senator Susan Collins*. Retrieved from http://www.collins.senate.gov/public/index.cfm/2015/8/senator-collins-continues-to-sound-the-alarm-on-the-urgent-need-to-bolster-cybersecurity.

Conway, M. (2008). Media, fear and the hyperreal: the construction of cyberterrorism as the ultimate threat to critical infrastructures. In M. Dunn Cavelty & K. S. Kristensen (Eds.), *Securing the 'Homeland': Critical Infrastructure, Risk and (In)Security* (pp. 109-129). London: Routledge. Retrieved from http://doras.dcu.ie/2142/1/2008-5.pdf.

Debrix, F. (2001). Cyberterror and media-induced fears: The production of emergency culture. *Strategies*, 14(1), 149-168.

Dunn Cavelty, M. (2008). *Cyber-Security and Threat Politics: U.S. Efforts to Secure the Information Age*. New York, NY: Routledge.

3 Pollan's dictum is 'Eat food. Not too much. Mostly plants.' (Pollan, 2008).

Dunn Cavelty, M., & Van Der Vlugt, R. A. (2015). A tale of two cities: Or how wrong metaphors lead to less security. *Georgetown Journal of International Affairs, Fall*, 21-29.

Elias, T. D. (1994, 2 January). Toffler: Computer attacks wave of future. *South Bend Tribune (Indiana)*, p. F10.

Furedi, F. (2009). Precautionary culture and the rise of possibilistic risk assessment. *Erasmus Law Review*, 2(2), 197-220.

Gallagher, R., & Greenwald, G. (2014, 12 March). How the NSA plans to infect 'millions' of computers with malware. *The Intercept*. Retrieved from https://firstlook.org/theintercept/2014/03/12/nsa-plans-infect-millions-computers-malware/.

Glassner, B. (1999). *The culture of fear: why Americans are afraid of the wrong things*. New York, NY: Basic Books.

Herath, T., & Rao, H. R. (2009). Protection motivation and deterrence: A framework for security policy compliance in organisations. *European Journal of Information Systems, 18*(2), 106-125.

Koppel, T. (2015). *Lights Out: A Cyberattack, A Nation Unprepared, Surviving the Aftermath*. New York: Crown.

Lawson, S. (2012, 16 October). Of cyber doom, dots, and distractions. *Forbes.com*. Retrieved from http://www.forbes.com/sites/seanlawson/2012/10/16/of-cyber-doom-dots-and-distractions/.

Lawson, S. (2013a). Beyond cyber-doom: Assessing the limits of hypothetical scenarios in the framing of cyber-threats. *Journal of Information Technology & Politics, 10*(1), 86-103.

Lawson, S. (2013b). Motivating cybersecurity: Assessing the status of critical infrastructure as an object of cyber threats. In C. Laing, A. Badii, & P. Vickers (Eds.), *Securing critical infrastructures and critical control systems: Approaches for threat protection* (pp. 168-189). Hershey, PA: IGI Global.

Ledbetter, S. (2015, 13 October). America's top fears 2015. *Chapman University Blog*. Retrieved from https://blogs.chapman.edu/wilkinson/2015/10/13/americas-top-fears-2015/.

Lee, D., Larose, R., & Rifon, N. (2008). Keeping our network safe: A model of online protection behaviour. *Behaviour & Information Technology, 27*(5), 445-454.

Lewis, J. A. (2010). The Cyber War Has Not Begun. *Center for Strategic and International Studies*. Retrieved from http://csis.org/files/publication/100311_TheCyberWarHasNotBegun.pdf.

Lyngaas, S. (2015, 23 February). NSA's Rogers makes the case for cyber norms. *FCW*. Retrieved from https://fcw.com/articles/2015/02/23/nsa-rogers-cyber-norms.aspx.

Maddux, J. E., & Rogers, R. W. (1983). Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of Experimental Social Psychology, 19*(5), 469-479. doi: 10.1016/0022-1031(83)90023-9.

Martinez, J. (2012, 31 October). Napolitano: Us financial institutions 'actively under attack' by hackers. *The Hill*. Retrieved from http://thehill.com/policy/technology/265167-napolitano-us-financial-institutions-qactively-under-attackq-by-hackers.

Meyer, D. (2010). Cyberwar could be worse than a tsunami. *ZDNet*. Retrieved from http://www.zdnet.com/news/cyberwar-could-be-worse-than-a-tsunami/462576.

Obama, B. (2015, 13 February). Remarks by the president at the cybersecurity and consumer protection summit. *Office of the Press Secretary, The White House*. Retrieved from https://www.whitehouse.gov/the-press-office/2015/02/13/remarks-president-cybersecurity-and-consumer-protection-summit.

Pagliery, J. (2015, 27 October). Senate overwhelmingly passes historic cybersecurity bill. *CNN*. Retrieved from http://money.cnn.com/2015/10/27/technology/cisa-cybersecurity-information-sharing-act/.

Panetta L (2012) Defending the National From Cyber Attacks. Presentation to Business Executives for National Security, New York, NY. 11 October.

Peters, G.-J. Y., Ruiter, R. A. C., & Kok, G. (2013). Threatening communication: A critical re-analysis and a revised meta-analytic test of fear appeal theory. *Health Psychology Review, 7*(sup1), S8-S31.

Pfau, M. (2007). Who's afraid of fear appeals? Contingency, courage, and deliberation in rhetorical theory and practice. *Philosophy and Rhetoric, 40*(2), 216-237.

Pfleeger, S. L., & Caputo, D. D. (2012). Leveraging behavioral science to mitigate cyber security risk. *Computers & Security, 31*(4), 597-611.

Pollan, M. (2008). *In defense of food: an eater's manifesto*. New York: Penguin Press.

Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change. *The Journal of Psychology, 91*(1), 93-114. doi: 10.1080/00223980.1975.9915803.

Schneier, B. (2014, 14 August). Quantum technology sold by cyberweapons arms manufacturers. *Schneier on Security*. Retrieved from https://www.schneier.com/blog/archives/2014/08/quantum_technol.html.

Schwartau, W. (1991). *Terminal Compromise*. Seminole, FL: Inter.Pact Press.

Seidman, R. (2015, February 22). List of how many homes each cable network is in as of February 2015. *TV by the Numbers*. Retrieved October 1, 2015, from http://tvbythenumbers.zap2it.com/2015/02/22/list-of-how-many-homes-each-cable-network-is-in-as-of-february-2015/366230/.

Singer PW (2016, 11 February) Cybersecurity and Cyberwar: What Everyone Needs to Know. Presentation to College of Wooster. Retrieved from http://www.wooster.edu/news/releases/2016/february/recap-singer/index.php.

Siponen, M., Adam, M., M., & Pahnila, S. (2014). Employees' adherence to information security policies: An exploratory field study. *Information & Management, 51*(2), 217-224.

Stohl, M. (2007). Cyber terrorism: A clear and present danger, the sum of all fears, breaking point or patriot games? *Crime, Law and Social Change, 46*(4-5), 223-238.

The Atlantic. (2010, 30 September). Fmr. Intelligence director: New cyberattack may be worse than 9/11. *The Atlantic*. Retrieved from http://www.theatlantic.com/politics/archive/2010/09/fmr-intelligence-director-new-cyberattack-may-be-worse-than-9-11/63849/.

Thierer, A. (2013). Technopanics, threat inflation, and the danger of an information technology precautionary principle. *Minn. JL Sci. & Tech., 14*, 309. Retrieved from http://conservancy.umn.edu/bitstream/11299/144225/1/Technopanics-by-Adam-Thierer-MN-Journal-Law-Science-Tech-Issue-14-1.pdf.

Valeriano, B. and Ryan C Maness. (2015). *Cyber War Versus Cyber Realities: Cyber Conflict in the International System*. London: Oxford University Press.

Walton, D. N. (2000). *Scare tactics: arguments that appeal to fear and threats*. Boston: Kluwer Academic Publishers.

Weimann, G. (2005). Cyberterrorism: The Sum of All Fears? *Studies in Conflict & Terrorism, 28*(2), 129-149. doi:10.1080/10576100590905110

Weimann, G. (2008). Cyber-terrorism: Are we barking at the wrong tree? *Harvard Asia Pacific Review, 9*(2), 41-46.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Determining Extremist Organisations' Likelihood of Conducting Cyber Attacks

Steve S. Sin
National Consortium for the Study of Terrorism and Responses to Terrorism (START)
University of Maryland
College Park, MD, USA
sinss@umd.edu

Elvis Asiamah*
National Consortium for the Study of Terrorism and Responses to Terrorism (START)
University of Maryland
College Park, MD, USA
easiamah@mix.wvu.edu

Laura A. Blackerby*
School of International Service
American University
Washington, D.C., USA
lauraablackerby@gmail.com

Rhyner Washburn*
The Graduate School: Cyber security
University of Maryland, Baltimore County
Baltimore, MD, USA
rhynerwashburn@gmail.com

Abstract: The 2007 cyber attacks against Estonia were a prime example of how hackers can operate within a nation's cyber domain. Looking at today's cyber landscape, it would seem that the tools needed for an offensive cyber operational capability are readily available to a properly motivated extremist organisation. Reports and articles about hacking and cyber security from think tanks and popular publications focused on technology and business tend to reinforce such a perception. In the nexus of cyber and physical and especially in the context terrorism, given the availability of offensive cyber capability, it is unclear why more extremist organisations are not engaging in offensive cyber operations. To investigate this anomaly this study employed a structured expert elicitation to identify a set of variables that would assist in determining an extremist organisation's likelihood of choosing to engage in cyber attacks. The results revealed that while there are points of convergence as well as extreme divergences in the assessment, *level of Internet presence, access to human resources*, and *human resource available (internally to the organisation)* were assessed to have the most explanatory power for determining an extremist organisation's likelihood for engaging in offensive cyber operations.

Keywords: *offensive cyber capabilities, cyber physical nexus, future threat, extremist organisation, organisational characteristics*

* Second through fourth authors listed in alphabetical order

Any opinions, findings, or conclusions presented here are solely the authors' and are not representative of the National Consortium for the Study of Terrorism and Responses to Terrorism (START) or the University of Maryland.

1. INTRODUCTION

The transformation seen in the cyber landscape over the past decade has been both a blessing and a curse. Humanity has benefited from this evolution, but the current cyber threat-scape has many officials worried about the potential of extremist organisations conducting offensive operations in the cyber domain. The tools needed for offensive cyber operations seem to be readily available to highly motivated individuals or extremist organisations. 'Criminal activities in cyberspace are increasingly facilitated by burgeoning black markets for both tools (e.g., exploit kits) and take (e.g., credit card information) …' (Ablon, Libicki and Golay, 2014). 'Just as you can use Taskrabbit to find someone to fix your kitchen cabinets, so too can you find a hacker to perform digital espionage or destruction. Hackers are now offering their services online to the highest bidder' (Weissman, 2015). In the nexus of cyber and physical and especially in the context of terrorism, given the availability of offensive cyber capability, it is unclear why more extremist organisations are not engaging in, or attempting to engage in, offensive cyber operations either as a standalone or as a part of a comprehensive attack methodology?

There has been quite a bit of discussion about terrorist organisations' use of the Internet for recruitment and propaganda purposes, and there has been, of late, concern about the degree of cyber threat posed by extremist groups such as Islamic State (see, for example, Perlroth, 2015). Despite burgeoning discussions and literature on offensive cyber activities and the future of cyber security, these discussions and the literature have all too often been restricted to the technical domain. The extant literature, especially, tends to overlook or only pay scant attention to the fact that there is an individual or a group of individuals making the decision as to whether or not if their organisation will engage in offensive cyber operations, or will develop such a capability.

Since there are no preceding studies that systematically explore this issue, or that examine the human or organisational aspects of offensive cyber operations in the context of terrorism, research is needed to determine what characteristics make an extremist organisation more likely to engage in offensive cyber operations. We decided to take a three-phased approach to examining the issue. In Phase I, we conducted a structured expert elicitation to identify potential explanatory variables against which to collect data in Phase II of the research. In Phase II, we will construct a violent extremist organisation dataset that includes the cyber relevant variables identified in Phase I to empirically determine the organisational characteristics that make violent extremist organisations more likely to engage in offensive cyber operations. Finally, using the results of Phase II, we will conduct further research to elucidate the possible answers to the overall research question in Phase III – an empirical examination of why only a limited number of extremist organisations have engaged in offensive cyber operations thus far, as well as a forecast of the potential future threat-scape.

This paper, a product of Phase I of the research, discusses the results of the expert elicitation and provides some conclusions and implications. The remainder of this paper is organised into four sections. Firstly, it presents a summary review of literature that focuses on why extremist organisations have not been conducting offensive cyber operations. Secondly, research design

and methodology are discussed. Thirdly, the results of the structured expert elicitation are explicated, and finally, conclusions and implications are presented.

2. SUMMARY REVIEW OF RELEVANT LITERATURE

The largest difference between an offensive cyber operation and a conventional offensive military or terrorist operation is the domain where these operations take place. Offensive cyber operations are executed in (and through) the cyber domain, while conventional offensive military or terrorist operations are performed in the physical domain. While this distinction may seem like a gross simplification, the fact of the matter is that this difference in operational domains is what makes offensive cyber operations one of the most challenging (perhaps even the most challenging) national security issues of today and the future.

Scholars suggest that as conventional counter-terrorism measures against extremist organisations continue to increase, more groups will turn to the Internet as the next frontier for conducting offensive operations. The high level of anonymity that an individual or a group can maintain while operating in cyberspace, making law enforcement efforts and interdiction very difficult, as well as the relative low cost of such operations, are thought to be the primary reasons for this potential strategic shift.

Most aspects of modern life rely on Internet use and this nearly ubiquitous connectivity amplifies the opportunities for malicious actors to engage in cyber attacks against nearly anyone or anything from anywhere in the world while remaining anonymous (Cox, 2015; M86 Security Labs, 2010; Pawlak and Wendling, 2013). Aside from the near ubiquitous connectivity, the ease of purchasing cyber weapons, the support from malware developers and the broader black hat community, the relatively low skill level needed to become a 'hacker', and the heuristic nature of hacking, have all been assessed as primary reasons that have contributed to the increase in malicious actors' operations in cyberspace (Greenberg, 2012; Fossi, et al., 2008; Goncharov, 2013; Fortinet, 2012; Pawlak and Wendling, 2013; Peterson, 2013; Zhang, Jha, and Raghunathan, 2014).

Despite the apparent advantages the cyber domain affords various state and non-state actors to conduct nefarious activities in anonymity, it would appear (at least based on open source information), that large extremist organisations such as the Haqqani Network, al-Qaeda, and the Islamic State[1] have had little or no interest in conducting offensive cyber operations in earnest. This seems to be an apparent incongruity; if the cyber domain does indeed afford such advantages, why have extremist organisations not invested in using it much more than they have thus far? The literature addressing this issue largely refers to the mismatch between the effects of cyber operation and the overall goals of the extremist organisation as the primary barrier for the extremist organisation's engagement in offensive cyber operations.

1 Islamic State may be exploring the potential of incorporating offensive cyber operations into its overall attack methodology. According to Caitlin Durkovich, Assistant Secretary for Infrastructure Protection of the U.S. Department of Homeland Security, the Islamic State has been attempting to 'hack' into the U.S. electric grid (Pagliery, 2015). George Osborne, Chancellor of the Exchequer of the U.K., also purported that Islamic State is attempting to launch cyber attacks aimed at Britain's airports, hospitals, and the electricity supply (Bloom, 2015). Other reports seem to indicate that Islamic State does indeed have some very rudimentary cyber capabilities, but it has not made cyber capabilities or cyber operations one of its core goals (Perlroth, 2015).

The extant literature, drawing on the extensive terrorism literature, calls attention to the three broad goals that need to be met in order for an attack to be considered a success for extremist organisations: 1) incitement of fear on a massive scale that will reverberate beyond the initial attack; 2) satisfaction of a strategic political or religious objective; and 3) visual destruction in order to propagate fear (Archer, 2014; Cox, 2015; Kenney, 2015). While an extremist organisation could achieve the first two goals described above through a cyber attack relatively easily, attaining a visual destruction is exceptionally difficult through the cyber domain. Furthermore, to date, cyber attacks have been primarily used to disrupt rather than destroy, one of the core elements of a terrorist attack (Cox 2015).

The reason why cyber attacks have been used primarily for disruption rather than for destruction thus far can be attributed to the design of the cyber weapons themselves, as well as to the perpetrators' preferred outcomes from cyber attacks. First, typical cyber weapons, with the primary exception being Stuxnet and its variants, do not generally cause physical destruction since they are primarily designed to infiltrate a system to gather intelligence. Second, in the context of cyber attacks, disruption of a Wi-Fi network, electric grid, or database firewall is far preferable to destroying it, since disruption allows the perpetrators to move through the system unchecked and leave a backdoor so that they can return to the system in the future (Kenney, 2015; M86 Security Labs, 2010; NTT Group, 2015; Greenberg, 2012; Peterson, 2013; Zhang, Jha, and Raghunathan, 2014). Ponagi, et al. (2012) has also suggested that cyber weapons have a specific second-order-effect of psychological function, be it to spread confusion; foment distrust within the users of the targeted system; deceive; or distract so that the attacker can move through the system freely to obtain their goal. The assessment of the extant literature is that if a cyber attack were used to destroy rather than disrupt, quite apart from the technical challenge, the attacker would lose all the benefits that could be gained through cyber disruption.

Other factors that could serve as barriers to the extremist organisations' use of offensive cyber operation are attribution, cost, and sustainability. A central tenant of conducting an offensive cyber operation is to make sure the attack cannot be traced back to the originator. Non-attribution is vital because this allows the perpetrators to work freely with relatively low fear of being discovered or interdicted (Archer, 2014; Cox, 2015). Since one of the extremist organisations' modus operandi is to claim responsibility for and exploit the results of an attack, non-attribution would be out of kilter with these organisations' operational (and perhaps strategic) goals. One exception to this would be hacktivist organisations. Hacktivist organisations make a point of claiming responsibility for their cyber exploits; however, they are careful to erase their digital footprint in the targeted systems so that digital forensics cannot be performed (Seebruck, 2015).

Cost is another barrier. While offensive cyber operations can be executed by an individual at relatively low cost, operations that are more complex will require greater resources and involve more than one operator. The procuring of equipment, encryption software, exploits, and personnel to manage the operation can all potentially generate an enormous overhead and draw unwanted attention (Goncharov, 2013; Greenberg, 2012; Fortinet, 2012; Fossi et al., 2008). Furthermore, if the objective of the cyber attack does not merit the cost, then the operation can implode. Cost, therefore, is assessed to be the reason why most offensive cyber operations

conducted by non-state actors today tend to be associated with achieving monetary gains (Goncharov, 2013; Greenberg, 2012; M86 Security Labs, 2010).

Sustainability (or lack thereof) can serve as a barrier to extremist organisations engaging in offensive cyber operations. Sustainability is what separates an ordinary hacker from an Advanced Persistent Threat (APT). Offensive cyber operations rarely take a few minutes to execute. Rather, they are time-consuming affairs that may or may not yield acceptable results. It takes months, sometimes years, to plan and execute a complex cyber operation with no guarantee of any return on investment. There are very few extremist organisations that can afford and maintain the commitment to see it through, with potentially no visible benefit, for an extremely narrowly focused objective (Seebruck, 2015; Finklea et al., 2015).

The literature clearly demonstrates the potential advantages and relative ease of engaging in offensive cyber operations for a motivated individual or extremist organisation. It also provides several substantial reasons that could deter extremist organisations from engaging in and incorporating offensive cyber operations into their attack portfolios; however, there are a few notable deficiencies in the literature. First, it only examines the extremist organisations' use of offensive cyber operations as a standalone cyber exploit. For example, it assesses disruption-focused cyber weapons and the non-attributive nature of the offensive cyber operations as potential deterrents. These assessments may absolutely be correct if one only considers cyber only operations. However, as soon as one moves to combined cyber and physical operations where cyber operations are designed to facilitate the physical operations of traditional kinetic terrorist attacks, disruption and non-attribution of the cyber domain may just be what an extremist organisation needs to successfully execute the kinetic attacks in the physical domain. Second, the literature currently does not take into consideration the wide-ranging variations between extremist organisations in a systematic manner. Since variances in organisational characteristics can contribute to organisational behaviour, a systematic understanding of organisational characteristics can contribute to an extremist organisation's higher likelihood of obtaining or developing offensive cyber capabilities as well as engaging in offensive cyber operations is necessary.

3. RESEARCH DESIGN AND METHODOLOGY

Since there are no preceding studies that systematically examine the human and organisational aspects of offensive cyber operations in the context of terrorism, a structured expert elicitation was conducted to identify potential explanatory variables against which to collect data in Phase II. The structured expert elicitation was designed to ascertain expert assessments on the relative importance of various organisational characteristics. The primary tool used for this structured expert elicitation was a survey, and the participants were all currently active members of the public sector, private sector, or academia working on issues relevant to cyber security.

The survey consisted of presenting the experts with a set of 22 independent variables, and asking them to rate each variable on a five-point ordinal scale. These variables were drawn from

the list of extremist organisational variables and profiles of existing extremist organisations developed by START[2] and maintained over the past decade. Table 1 provides the list of all 22 organisational factors presented to the experts. The variables were grouped into four distinct factors based on their characteristics: organisational, resource, sophistication, and prior activities.

We recruited two groups of experts to serve as potential participants – those with technical backgrounds in subjects such as computer science or computer engineering, and those with non-technical backgrounds including policy makers, policy analysts, and academics. Once the potential participants were identified, the survey with detailed instructions was sent to them by email. All participants received the same survey. The survey response rate was 18.67%.

The collected data was coded into a format where statistical analysis could be performed to identify the variables that the participants assessed as having the most probabilistic explanatory power for determining an extremist organisation's likelihood of choosing to engage in offensive cyber operations. Table 2 provides a full list of participant groupings used in the analyses of the survey results, and the findings from the analyses are discussed in the next section.

4. FINDINGS

A. Overview

The data collected revealed there are several points of convergence, as well as divergence in the participants' assessment of organisational characteristics that could be significant in forecasting whether an extremist organisation is more or less likely to incorporate offensive cyber operations into its attack portfolio. Overall, there was a general consensus among the experts that *level of Internet presence, access to human resources*, and *human resource available* (internally to the organisation) are the variables with the most probabilistic explanatory power. The analysis of the survey results (see Table 3) found that the *level of Internet presence* of an organisation was the most important variable, while the statistical analysis of the results (see Table 4) revealed *access to necessary human resources* as the most powerful explanatory variable. Although *access to necessary human resources* was not found to be rated as the most important variable, it was in the top five variables (see Table 3), demonstrating that a degree of congruence does exist between the survey and statistical analyses results. The remainder of this section discusses the results of the analyses in further detail, beginning with the overall survey results.

B. Overall analysis

1) Results of survey analyses

The survey analysis found that the *level of Internet presence* of an organisation was rated as the most important variable by the participants. Their reasoning, according to their responses to the open ended questions in the survey, was that even if the organisation did not currently does not

2 Established in 2005, the National Consortium for the Study of Terrorism and Responses to Terrorism (START) is a research and education centre based at the University of Maryland and comprised of an international network of scholars committed to the scientific study of the causes and human consequences of terrorism in the United States and around the world. As a U.S. Department of Homeland Security Centre of Excellence, START supports the research efforts of leading social scientists at more than 50 academic and research institutions, each of whom is conducting original investigations into fundamental questions about terrorism.

have offensive cyber capabilities or had not engaged in offensive cyber operations, the fact that an organisation has a favourable disposition towards conducting portions of its operations in the cyber domain means it would be more comfortable with the idea of offensive cyber operations, and be more likely to consider incorporating offensive cyber operations into its attack portfolio, either as a standalone capability or as a part of a comprehensive cyber and physical attack package (Sin, et al., 2015). The participants also assessed that organisation's *leadership, access to human resources*, and *human resources available* (internal to the organisation) are top organisational characteristics (see Table 3).

Despite the fact that some consensus can be found in the participants' responses, there were some significant divergences in the ratings as well. For example, participants with technical backgrounds and those working in the public sector, rated *financial resources* available to the organisation as one of the top variables, while those with non-technical backgrounds and those who works in the private sector and academia included the organisation's *penchant for innovation* and *propensity for violence* in the list (see Table 3). *Access to necessary human resources* – the most significant variable in the statistical analyses – was assessed as one of the top five variables by four of the six participant groups. Figures 1 to 6 show the analyses of the participants' ratings of all variables. An interesting result shown in these figures is that the participants rated variables such as *ideology, organisational structure, organisational size*, and *network connections and access* – all factors found to be extremely significant in determining organisational behaviour and lethality in terrorism literature (Asal, Ackerman, and Rethemeyer, 2012; Asal and Rethemeyer, 2008; Gill, et al., 2014) – to be less important factors in forecasting an extremist organisation's likelihood of engaging in offensive cyber operations.

2) Results of statistical analysis

Statistical analysis conducted on the survey data to determine which variables the participants rated as having the highest explanatory power, found *access to necessary human resources* to be most significant. *Propensity for violence* and *human resources available* were found to be the second and third most significant respectively (see Table 4). All three variables were also found to be significant in the survey analyses, and all of the variables found to be significant in the statistical analysis are included in the list of top five most significant variables ascertained through the survey analysis (see Table 3).

The combined results of the statistical and survey analyses indicate there is a general consensus among the experts that *access to necessary human resources; propensity for violence; human resources available; level of internet presence; leadership characteristics;* and *penchant for innovation* are the most significant characteristics with explanatory power for predicting an organisation's likelihood of engaging in offensive cyber operations in the future.

C. Group analyses

A series of survey analyses were performed on five distinct groups of participants (see Table 2 for a list of groupings) to obtain a more nuanced understanding of the survey responses. The analyses revealed some very interesting within- and between-group trends. In general, a comparison of results between technical and non-technical groups yielded a very different picture than a comparison of results across the public sector, private sector, and academic

groups. Different groups, independently of each other, also exhibited clear and distinct trends of thought. Examining all of the seemingly disparate results, we were able to extrapolate that *Internet presence, leadership,* and *penchant for innovation* were considered to have the most probabilistic explanatory power. A more detailed comparison is discussed below.

1) Technical – non-technical group analysis

Overall, the technical group exhibited a higher within-group consensus on the importance and non-importance of variables than the non-technical group. The technical group tended to place relatively higher importance on the variables belonging to the *prior activities* factor such as *prior terrorist activities* and *propensity for violence*. Furthermore, the group placed a high value on the usability of resources such as *access to human resources*. The group also tended to assess variables relevant to organisational, membership, and leadership characteristics as being important, but to a much lesser degree than the organisation's prior activities and resources available. By contrast, the non-technical group placed the highest importance on the variables belonging to the *sophistication* factor, favouring the technical aspects of the organisation's characteristics, while assessing the variables belonging to other factors as less important or not important at all.

The analyses clearly showed a divergence in the perspectives on *organisational* factor between the two groups. Overall, the non-technical group assessed *organisational size* and *membership cognitive distortion* as not being important at all, while assessing *membership composition* as being almost neutral. The technical group, by contrast, assessed these factors as being more important compared to the non-technical group. Additionally, the technical group showed a tendency to place greater importance on leadership related variables than the non-technical group.

The technical group showed a strong consensus on variables belonging to *resource factors*. Of note, the two groups were shown to have an opposing view on *network access*, with the technical group assessing it not to be important and the non-technical group assessing it otherwise, albeit at a minimum importance. On other variables belonging to this factor, however, the two groups showed general agreement, with the technical group placing a much higher importance on them than the non-technical group.

The two groups showed a consensus in assessment for the variables belonging to the *sophistication* factor. Although there were some differences in the degree of importance placed on the variables between the two groups, they generally viewed all variables belonging to this factor as not being important with the exception of *Internet presence*, which was assessed as highly important by both groups.

2) Public sector – private sector – academic group analysis

Analyses of the public sector, private sector, and academic groups revealed that the public sector group tended to give high marks to resource and technical variables, while giving relatively lower marks to organisational, organisational composition, and leadership variables. Although the group assessed *Internet presence* and human and financial resources to be of high importance, it did not assess *network access* as important. The analysis of the public sector

group revealed that it perceived the decision of an extremist organisation to engage in offensive cyber operations is primarily a function of human, materiel, and financial resource availability.

The private sector group viewed *penchant for innovation* and *network access* to be more important than *Internet presence*, while assessing human and financial resource variables as relatively unimportant, a distinct divergence from the public sector group. The private sector group was the only group to assess *ideology* as being important among all the groups. The group also acknowledged the importance of the leadership variables, but assessed the relationship between variables relevant to organisational structure, size, and membership and the extremist organisation's likelihood to engage in offensive cyber operations as being quite low. The analysis of the private sector group revealed that this group perceived the decision of an extremist organisation to engage in offensive cyber operations as primarily a function of innovative ideas and having access to a network that could provide a support infrastructure.

The academic group, departing from the other two groups, placed low importance on the variables related to technical, historical, human resources, financial, and network aspects of the organisational characteristics. Compared to the other groups, it placed the lowest value on the *ideology* of the organisation and the highest on the *leadership* characteristics. The group also assessed *organisational cohesion* and *member composition* as being important, while *organisational structure* and *organisational size* were assessed as not. The results revealed that the academic group perceived leadership characteristics and the relationship between the leaders as key factors that could predict which extremist organisations was mostly likely to engage in offensive cyber operations.

Some interesting divergences were observed in the groups' assessments of the *organisational* factor variables. The private sector group was the only group that assessed *ideology* as being important, and the other two assessed it as strongly negative (unimportant), and the academic group was the only one to assess *membership composition* as being important. Additionally, the public group assessed *leadership* as not being very important while the other two groups assessed it as being important. The academic group in particular assessed *leadership* to be extremely important. In fact, the academic group was the only group that assessed all leadership related variables (with the exception of *leadership cognitive distortion*) as being very important in predicting an extremist organisation's likelihood of engaging in offensive cyber operations.

Consensus was apparent amongst the three groups on the importance of *organisational structure* and *organisational size* as not being very important, with the private sector group assessing them most negatively (unimportant) compared to the other two groups. The consensus amongst the groups diverged once again in their assessment of *organisational cohesion* with both public sector and academic groups assessing it to be important and the private sector group assessing it to be unimportant.

5. CONCLUSION

As the first phase to understanding why extremist organisations are not incorporating offensive cyber operations into their attack portfolios, this research sought to identify potential explanatory variables that can be incorporated into the development of a violent extremist organisation dataset, which includes cyber relevant variables. Once developed, the dataset will be used to empirically determine the organisational characteristics that make violent extremist organisations more likely to engage in offensive cyber operations. The results of this determination can then be used to explore further why they do not appear to be engaging in offensive cyber operations as much as they theoretically could.

In this phase of the research, a structured expert elicitation was conducted through a survey where the participants assessed the importance of 22 variables. The results of the survey and the statistical analyses found the participants assessed organisations' *level of Internet presence, access to human resources*, and *human resources available* as the variables most likely to predict the likelihood of a violent extremist organisation's decision to incorporate offensive cyber operations into its portfolio.

This research was also able to identify some important points of consensus and divergence that exist between the various expert groups. The consensus and divergence in assessed importance of variables were most significant along the participants' background and occupation. Analysis of each group showed the participants tended to have internal consistency within their assigned groups, but the groups showed clear between-group divergence. Second, the group divergences based on participant backgrounds were much weaker than the divergences observed among occupation-based groups. The degree of variance was much higher among the public sector, private sector, and academic groups than between the technical and non-technical groups. Finally, each group had a varying degree of internal group consistency. Between the technical and non-technical groups, the technical group exhibited a much higher internal group consistency than the non-technical group. Among the public sector, private sector, and academic groups, the public sector group presented the highest degree of internal group consistency, followed by the private sector group. The academic group exhibited the least internal group consistency. These results not only illuminate the impact that occupation has on one's world view in terms of perspectives and the degree of conformity, but they also expose a lack of clear understanding among the groups about each other's perspectives.

Another interesting trend observed during this research was the similarity in the results between the technical and the public sector groups. While some differences did exist between the two groups, they shared similar assessments of variables, both in direction (important or not important) and magnitude. Given that the public sector group was divided almost equally between technical and non-technical participants (53% technical and 47% non-technical), this trend suggests that: 1) unlike the results of the overall technical – non-technical comparisons, a much higher degree of consensus exists between the technical and non-technical groups within the public sector; and 2) the non-technical experts and practitioners in the public sector appear to have responded to the survey closer to their technical counterparts than the non-technical

participants in the other two sectors. While the current research cannot ascertain the exact cause of these differences, it could be that non-technical experts and practitioners perhaps have more routine exposure to technical information and experts during the course of their work. This is only speculation, however, and is a topic worth examining in future research.

A. Implications

Although this research is only the first phase of a three-phase research endeavour, the results have yielded several issues that require some consideration. First, the research confirmed that there is no systematic research being conducted today to determine why some violent extremist organisations decide to engage in offensive cyber operations, while others do not. Neither does it explain why few violent extremist organisations engage in offensive cyber operations despite the purported ease of acquiring the tools necessary to carry out such an operation. The research confirmed our impression that there is an apparent lack of standardised indicators that can be used to identify which violent extremist organisations are more likely to engage in offensive cyber operations. They also indicate there is a clear divide in focus and opinion and no evidence of robust communication between the various disciplines and sectors involved in cyber security. Finally, the results suggest that there may be a higher level of technical/non-technical expert/practitioner consensus than in other sectors examined, and follow-on research examining the convergences and divergences among experts and practitioners in different sectors is warranted.

These implications highlight the pressing need for the experts and practitioners of cyber security of various backgrounds and occupational areas to bridge the fundamental divides that exist among them through communication and education. They also call attention to a need for a broader cyber security research agenda that is multilateral, multidiscipline, and multimethod, and is designed to incorporate stakeholders from all sectors of society to address the challenges of the future cyber threat-scape.

B. Further research

Narrowing the focus back to the current research, we have developed the violent extremist organisation dataset that includes cyber-relevant variables by appending the cyber variables to the Big, Allied and Dangerous (BAAD) dataset (Asal, Rethemeyer, and Anderson, 2011), creating an extremist organisation cyber attack dataset. We are currently engaged in data collection following which, a series of follow-on research projects can be conducted to further explicate extremist organisations' likelihood of engaging in offensive cyber operations. Results from this research will allow an empirically based and more nuanced understanding of the relationships between terrorism, cyber, and extremist organisational behaviour.

ACKNOWLEDGMENTS

We would like to thank all of the subject matter experts who participated in the survey for this phase of the research. The team would also like to thank Dr Gary Ackerman, Director of Unconventional Weapons & Technology Division of START, University of Maryland, and Dr H. Sophia Sin, Research Associate of the Centre for Policy Research, State University of New York at Albany, for their support and encouragement throughout the course of this research.

REFERENCES

Ablon, Lillian, Martin C. Libicki, and Andrea A. Golay. 2014. *Markets for Cybercrime Tools and Stolen Data: Hackers' Bazaar*. National Security Research Division, RAND Corporation, Santa Monica, CA: RAND Corporation, 85.

Archer, Emrald M. 2014. 'Crossing the Rubicon: Understanding Cyber Terrorism in the European Context.' *The European Legacy* 19 (5): 606-621.

Asal, Victor, Gary Ackerman, and R. Karl Rethemeyer. 2012. 'Connections Can Be Toxic: Terrorist Organisational Factors and the Pursuit of CBRN Weapons.' *Studies in Terrorism and Conflict* 35: 229-254.

Asal, Victor, and R. Karl Rethemeyer. 2008. 'The Nature of the Beast: Terrorist Organisational Characteristics and Organisational Lethality.' *Journal of Politics* 70 (2): 437-449.

Asal, Victor, R. Karl Rethemeyer, and Ian Anderson. 2013. 'Big Allied and Dangerous (BAAD) Database, 1998-2012'.

Bloom, Dan. 2015. 'ISIS Trying to Launch Deadly Cyber Attack on Airports, Hospitals, and National Grid, Warns George Osborne.' *The Mirror*. November 17. Accessed December 30, 2015. http://www.mirror.co.uk/news/uk-news/isis-trying-launch-deadly-cyber 6845517.

Cox, Christopher. 2015. 'Cyber Capabilities and Intent of Terrorist Forces.' *Information Security Journal* 24 (1-3): 31-38.

Finklea, Kristin, Michelle Christensen, Eric Fischer, Susan Lawrence, and Catherine Theohary. 2015. 'Cyber Intrusion into U.S. Office of Personnel Management: In Brief.' CRS Report No. R44111, Congressional Research Service.

Fortinet. 2012. 'Cybercriminals Today Mirror Legitimate Business Processes.' 2013 Cybercrime Report, Fortinet, Inc.

Fossi, Marc, Eric Johnson, Dean Turner, Trevor Mack, Joseph Blackbird, David McKinney, Mo King Low, Teo Adams, Marika Pauls Laucht, and Jesse Gough. 2008. 'Symantec Report on the Underground Economy July 07-Jun 08.' Symantec Corporation.

Gill, Paul, Jeongyoon Lee, R. Karl Rethemeyer, John Horgan, and Victor Asal. 2014. 'Lethal Connections: The Determinants of Network Connections in the Provisional Irish Republican Army 1970 – 1998.' *International Interactions: Empirical and Theoretical Research in International Relations* 40 (1): 52-78.

Goncharov, Max. 2013. *Russian Underground* 101. Trend Micro Incorporated.

Greenberg, Andy. 2012. *Shopping for Zero Days: A Price List for Hackers' Secret Software Exploits*. March 23. Accessed December 28, 2015. http://www.forbes.com/sites/andygreenberg/2012/03/23/shopping-for-zero-days-an-price-list-for-hackers-secret-software-exploits/.

Kenney, Michael. 2015. 'Cyber Terrorism in a Post-Stuxnet World.' Orbis 59 (1): 111-128.

M86 Security Labs. 2010. *Web Exploits: There's an App for That*. M86 Security.

NTT Group. 2015. *Exploit Kits: Development and Operation Lifecycles*. Global Threat Intelligence Report, NTT Innovation Institute, LLC.

Pagliery, Jose. 2015. 'ISIS is Attacking the U.S. Energy Grid (and Failing).' *CNN Money*. October 16. Accessed December 30, 2015. http://money.cnn.com/2015/10/15/technology/isis-energy-grid/.

Pawlak, Patryk, and Cecile Wendling. 2013. 'Trends in Cyberspace: Can Governments Keep Up?' *Environment Systems & Decisions* 33 (4): 536-543.

Perlroth, Nicole. 2015. 'Security Experts and Officials Diverge on ISIS as Hacking Threat.' *The New York Times*. December 24. Accessed December 29, 2015. http://www.nytimes.com/2015/12/25/technology/security-experts-and-officials-diverge-on-isis-as-hacking-threat.html?_r=0.

Peterson, Dale. 2013. 'Offensive Cyber Weapons: Construction, Development, and Employment.' *Journal of Strategic Studies* 36 (1): 120-124.

Ponangi, Preethi Vinayak, Phani Kidambi, Dhananjai Rao, Narasimha Edala, Mary Fendley, Michael W. Haas, and S. Narayanan. 2012. 'On the Offense: Using Cyber Weapons to Influence Cognitive Behaviour.' *International Journal of Cyber Society & Education* 5 (2): 127-150.

Seebruck, Ryan. 2015. 'A Typology of Hackers: Classifying Cyber Malfeasance Using a Weighted Arc Circumplex Model.' *The Journal of Digital Investigation* (14) 3: 36-45.

Sin, Steve S., Elvis Asiamah, Laura A. Blackerby, and Rhyner Washburn. 2015. *Survey Conducted for CyCon 2016 Research*. College Park, MD, November 1.

Weissman, Cale Guthrie. 2015. 'It's Becoming Easier and Easier to 'Rent a Hacker'.' *Business Insider*. May 12. Accessed December 29, 2015. http://www.businessinsider.com/the-hacker-for-hire-market-is-growing-2015-5.

Zhang, Meng, Niraj Jha, and Anand Raghunathan. 2014. 'A Defense Framework Against Malware and Vulnerability Exploits.' *International Journal of Information Security* 13 (5): 439-452.

TABLES AND FIGURES

FIGURE 1: RATING OF ORGANISATIONAL FACTORS FOR FORECASTING LIKELIHOOD OF OFFENSIVE CYBER OPERATIONS (ALL PARTICIPANTS)

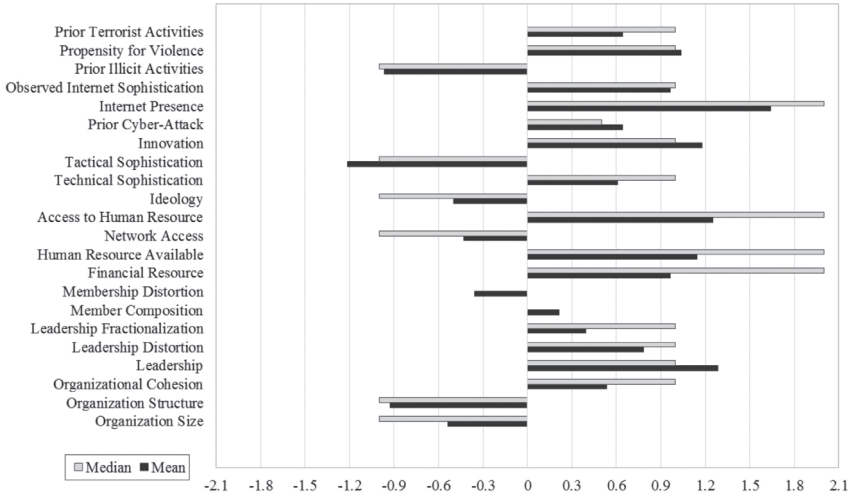


FIGURE 2: RATING OF ORGANISATIONAL FACTORS FOR FORECASTING LIKELIHOOD OF OFFENSIVE CYBER OPERATIONS (TECHNICAL PARTICIPANTS)

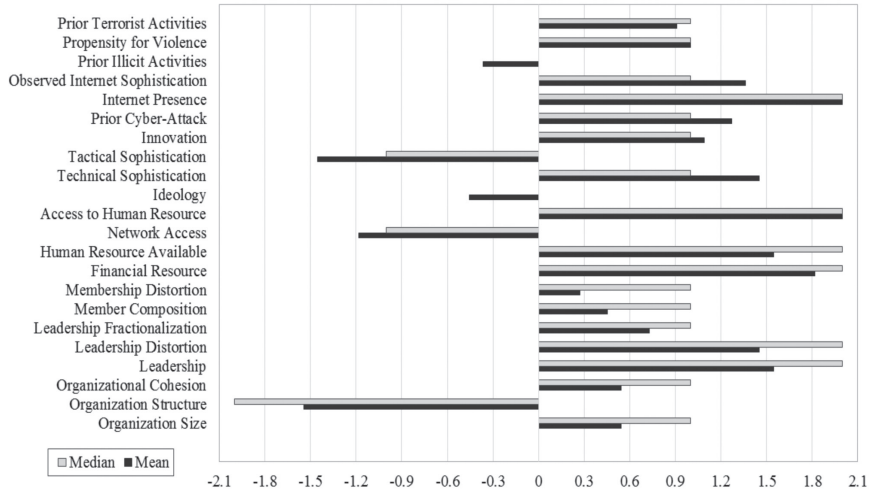


FIGURE 3: RATING OF ORGANISATIONAL FACTORS FOR FORECASTING LIKELIHOOD OF OFFENSIVE CYBER OPERATIONS (NON-TECHNICAL PARTICIPANTS)

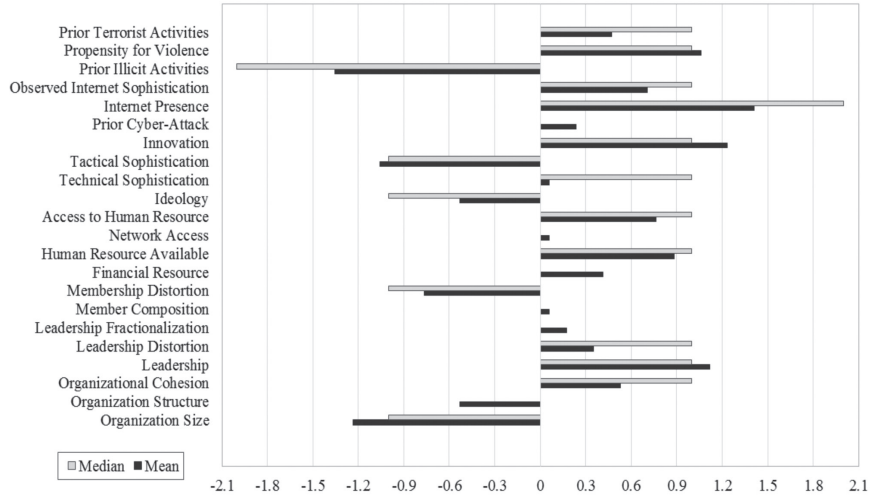


FIGURE 4: RATING OF ORGANISATIONAL FACTORS FOR FORECASTING LIKELIHOOD OF OFFENSIVE CYBER OPERATIONS (PUBLIC SECTOR PARTICIPANTS)

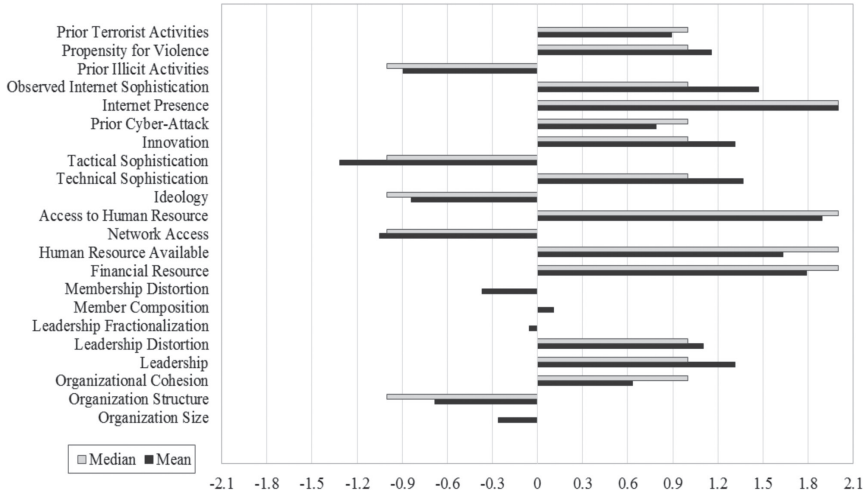


FIGURE 5: RATING OF ORGANISATIONAL FACTORS FOR FORECASTING LIKELIHOOD OF OFFENSIVE CYBER OPERATIONS (PRIVATE SECTOR PARTICIPANTS)

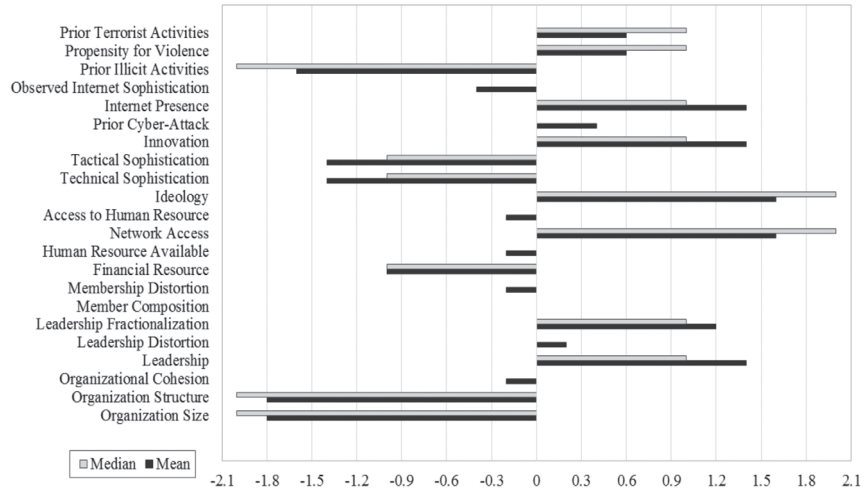


FIGURE 6: RATING OF ORGANISATIONAL FACTORS FOR FORECASTING LIKELIHOOD OF OFFENSIVE CYBER OPERATIONS (ACADEMIC PARTICIPANTS)

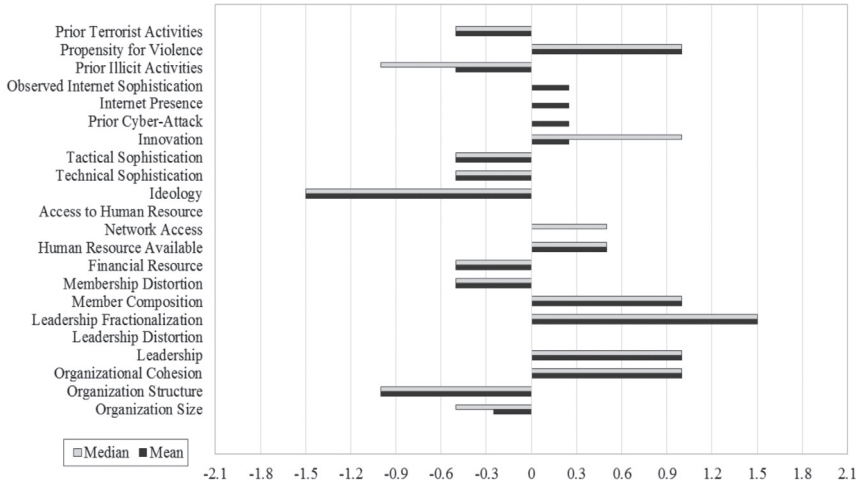


TABLE 1: ORGANISATIONAL CHARACTERISTICS INCLUDED IN THE SURVEY

Organisational Factors	Resource Factors	Sophistication Factors	Prior Activities Factors
Organisation Size	Financial Resources Available	Technical Sophistication	Prior History of Illicit Activities
Organisational Structure	Human Resources Available (Internal)	Tactical Sophistication	Propensity for (Prior History of) Violence
Organisational Cohesiveness	Network Connections & Access	Penchant for Innovation	Prior History of Cyber attack
Ideology	Access to Necessary Human Resources	Level of Internet Presence	Prior History of Terrorist Activities
Leadership		Observed Internet Sophistication	
Leadership Cognitive Distortion			
Leadership Fractionalisation			
Membership Composition			
Membership Cognitive Distortion			

TABLE 2: SURVEY PARTICIPANT GROUPINGS

Grouping 1	All Participants		
Grouping 2	Technical Participants	Non-Technical Participants	
Grouping 3	Public Sector Participants	Private Sector Participants	Academic Participants

TABLE 3: TOP FIVE ORGANISATIONAL CHARACTERISTICS RATING BY PARTICIPANT GROUPS (DESCRIPTIVE STATISTICS)

	Participants					
Rating	**All**	**Technical**	**Non-Technical**	**Public Sector**	**Private Sector**	**Academic**
1	Level of Internet Presence	Level of Internet Presence* Access to Human Resources*	Level of Internet Presence	Level of Internet Presence	Network Access* Ideology*	Leadership Fractionalisation
2	Leadership		Penchant for Innovation	Access to Human Resources		Leadership@ Organisational Cohesion@ Member Composition@ Propensity for Violence@
3	Access to Human Resources	Financial Resources	Leadership	Financial Resources	Level of Internet Presence# Penchant for Innovation# Leadership#	
4	Penchant for Innovation	Leadership^ Human Resources Available^	Propensity for Violence	Observed Internet Sophistication		
5	Human Resources Internally Available		Human Resources Available	Technical Sophistication		

* Tied for First
@ Tied for Second
\# Tied for Third
^ Tied for Fourth

TABLE 4: REGRESSION RESULTS (ALL PARTICIPANTS)

Dependent Variable: Organisation Engages in Offensive Cyber Operations

Independent Variable	Coefficient	Standard Error	T Score
Organisational Size	-.0116685	.083372	-0.14
Organisational Structure	.0254537	.0570688	0.45
Organisational Cohesion	.1014259	.0681252	1.49
Leadership	.0019509	.0570208	0.03
Leadership Cognitive Distortion	-.0052943	.0262764	-0.20
Leadership Fractionalisation	.0246953	.0506852	0.49
Membership Composition	-.00218	.0384258	-0.06
Membership Cognitive Distortion	.0183884	.0349441	0.53
Financial Resources Available	-.0743241	.0565408	-1.31
Human Resources Available (Internal)	.0487971	.0175308	2.78**
Network Connections and Access	.0532817	.0328357	1.62
Access to Necessary Human Resources	.0746465	.0237544	3.14**
Ideology	.025081	.0423652	0.59
Technical Sophistication	.0157153	.0607535	0.26
Tactical Sophistication	-.0097967	.0417712	-0.23
Penchant for Innovation	.0050692	.0239459	0.21
Prior History of Cyber attack	-.0040141	.0418024	-0.10
Level of Internet Presence	-.0069046	.0513136	-0.13
Level of Internet Sophistication	.0064245	.0448341	0.14
Prior History of Illicit Activities	.0306632	.0347698	0.88
Propensity for (Prior History of) Violence	.0630513	.015214	4.14***
Prior History of Terrorist Activities	.0383541	.0343559	1.12
Constant	.8387801	.2256389	3.72***

R-squared: 0.9977
All significance tests are two-tailed: *p<0.05; **p<0.01' ***p<0.001
Robust Standard Error used for Analysis

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

The Social Side of 'Cyber Power'? Social Media and Cyber Operations

Drew Herrick
Department of Political Science
George Washington University
Washington D.C., USA

Abstract: Evaluating an actor's 'cyber power' is an inherently complex problem involving a laundry list of military, normative, and technical variations. However, one important but under-theorised factor is the relationship between *military* social media operations and cyber operations. Policymakers, journalists, and even some academics often treat social media activity as a proxy variable for an actor's latent technical proficiency and even cyber capability, in other words, its cyber power. Actors that are extremely successful at engaging in social media activities are assumed to be technically proficient and even capable of engaging in cyber operations. This paper argues that an actor's social media use is a poor proxy for its technical and cyber security competency. In fact, under certain conditions social media activity may actually magnify the vulnerability of that actor. This paper synthesises cross-disciplinary research from strategic studies, political science, and technologists to develop a theoretical framework for better understanding the role of social media in cyber operations. It outlines the similarities and differences between social media and cyber security, and categorises different military social media operations into three types: information-gathering (IGMO), defensive social media operations (DeSMO), and offensive social media operations (OSMO).

Keywords: *future threats, situational awareness, data/information as power, international norms and governance, information operations*

1. INTRODUCTION

Correctly measuring an actor's offensive and defensive cyber capabilities or its aggregate 'cyber power' is an important goal for both policymakers and academics.[1] Knowing an actor's *true* capabilities affects not only expectations of success or failure on the battlefield, but also peacetime bargaining situations, escalation dynamics, balancing, deterrence, and even the durability of international norms.[2]

Unfortunately, evaluating an actor's cyber capabilities ex ante is extremely difficult for at least four reasons. First, technology in cyberspace is inherently dual use.[3] Even under ideal conditions, an accurate assessment of an actor's technological capabilities does not sufficiently reveal whether those capabilities are offensive or defensive in nature, assuming such distinctions even make sense.[4] Second, traditional assessment tools such as counting troops and materiel do not work well in a cyber context. Physical instantiations of cyber capabilities are rare.[5] Even attempts to examine an actor's ratio of successful to unsuccessful cyber operations are riddled with severe data limitations and selection bias issues.[6]

Third, in some cases there may be public financial or personnel disclosures that reveal how much money is being allocated to distinct operational areas, or how many people are working

1 This paper uses 'cyber capabilities' and 'cyber power' interchangeably. Other non-military elements that may or may not be part of an actor's aggregate cyber power such as commercial sector variables and Internet governance are bracketed. For a good overview on conceptualizing cyber power see Joseph S. Nye Jr., 'Cyber Power' (Cambridge: Belfer Center for Science and International Affairs, May 2010), http://belfercenter.ksg.harvard.edu/files/cyber-power.pdf; Joseph S. Nye Jr, *The Future of Power* (New York: PublicAffairs, 2011); 'Cyber Power Index. Findings and Methodology' (Economist Intelligence Unit, 2011), http://www.boozallen.com/content/dam/boozallen/media/file/Cyber_Power_Index_Findings_and_Methodology.pdf; Adam Segal, *The Hacked World Order: How Nations Fight, Trade, Maneuver, and Manipulate in the Digital Age* (New York: PublicAffairs, 2016).

2 For kinetic examples see, James D. Fearon, 'Rationalist Explanations for War,' *International Organization* 49, no. 3 (1995): 379–414; Robert Powell, 'Bargaining Theory and International Conflict,' *Annual Review of Political Science* 5, no. 1 (2002): 1–30, doi:10.1146/annurev.polisci.5.092601.141138; Stephen Biddle, *Military Power: Explaining Victory and Defeat in Modern Battle* (Princeton, N.J.; Oxford: Princeton University Press, 2006); Keir A. Lieber, *War and the Engineers: The Primacy of Politics over Technology* (Ithaca; London: Cornell University Press, 2008); Charles L. Glaser, *Rational Theory of International Politics: The Logic of Competition and Cooperation* (Princeton University Press, 2010).

3 Trey Herr and Paul Rosenzweig, 'Cyber Weapons & Export Control: Incorporating Dual Use with the PrEP Model,' *Journal of National Security Law & Policy* 8 (2015), http://jnslp.com/2015/10/23/cyber-weapons-export-control-incorporating-dual-use-with-the-prep-model/.

4 Keir A. Lieber, 'Mission Impossible: Measuring the Offense-Defense Balance with Military Net Assessment,' *Security Studies* 20, no. 3 (2011): 451–59; Stephen Biddle, 'Rebuilding the Foundations of Offense-Defense Theory,' *Journal of Politics* 63, no. 3 (August 1, 2001): 741–74, doi:10.1111/0022-3816.00086.

5 Jon R. Lindsay, 'Stuxnet and the Limits of Cyber Warfare,' *Security Studies* 22, no. 3 (July 1, 2013): 365–404, doi:10.1080/09636412.2013.816122; Erik Gartzke and Jon R. Lindsay, 'Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace,' *Security Studies* 24, no. 2 (April 3, 2015): 316–48, doi:10.1080/09636412.2015.1038188; Brandon Valeriano and Ryan C. Maness, 'The Fog of Cyberwar,' *Foreign Affairs*, December 21, 2015, https://www.foreignaffairs.com/articles/2012-11-21/fog-cyberwar; Drew Herrick and Trey Herr, 'Combating Complexity: Offensive Cyber Capabilities and Integrated Warfighting' (International Studies Association, Atlanta, GA, 2016).

6 Attribution problems and the covert nature of cyber operations make data collection extremely difficult. Both sides of a conflict may have incentives to strategically not report or misreport an incident. In the most advanced cases, successful cyber operations may not even be known let alone publically reported.

in a specific area.[7] Unfortunately, even this is an unreliable metric since cyber power is better framed as a function of actor skill and time, not of allocated raw resources.[8] Throwing large amounts of money or people at a problem may or may not be sufficient to close a large skill gap or neutralise first mover advantages. More importantly, states have strong incentives to misrepresent their capabilities or even take credit (or not take credit) for past successful operations regardless of their actual participation or true capability.

Finally, power is relational and therefore even having a static measure of one actor's cyber capabilities is not particularly helpful. Instead, observers need to have a more dynamic and relational measure of multiple actors' capabilities over time. The core problem is that even under ideal conditions there is still a large degree of uncertainty that afflicts operational planning and peacetime bargaining situations. In many situations, having a poor assessment of an actor's cyber power may be just as damaging or even more so than having no prior knowledge.[9] For example, assessments of another actor's power that are too low may incentivise various states to engage in risky or escalatory behaviour that they otherwise should avoid. Similarly, assessments that are on the high side may incentivise states to select out of conflicts that in reality they are well placed to win.

One possible solution to the uncertainty problem is for states to leverage their well-established intelligence apparatus to gather information and narrow the gap. However, even for advanced states it is unlikely that espionage can completely close the uncertainty gap. Regardless, the key issue is that the public and the cyber security research community face a large data collection problem and are forced to rely on declassified documents, interviews, and open source alternatives.[10] Therefore, finding a reliable set of public and directly observable proxy variables to measure an actor's latent cyber capabilities is critical. One potential variable that is repeatedly referenced by policymakers, journalists and even some academics is 'advanced social media use' by so-called 'keyboard warriors' or 'cyber-jihadis'[11] Unfortunately as will be

7 Aliya Sternstein, 'The Military's Cybersecurity Budget in 4 Charts,' *Defense One*, March 16, 2015, http://www.defenseone.com/management/2015/03/militarys-cybersecurity-budget-4-charts/107679/; 'China Creates 3 New Army Units to Modernize Military,' *The Washington Post*, January 1, 2016, https://www.washingtonpost.com/world/asia_pacific/china-creates-3-new-army-units-to-modernize-military/2016/01/01/33648432-b10a-11e5-b281-43c0b56f61fa_story.html.

8 Drew Herrick and Trey Herr, 'Combating Complexity.'

9 See for example, Randall L. Schweller, *Unanswered Threats: Political Constraints on the Balance of Power* (Princeton University Press, 2006); Aaron L. Friedberg, *The Weary Titan: Britain and the Experience of Relative Decline, 1895-1905* (Princeton University Press, 1988).

10 For example, see Kim Zetter, 'Security Manual Reveals the OPSEC Advice ISIS Gives Recruits,' *WIRED*, November 19, 2015, http://www.wired.com/2015/11/isis-opsec-encryption-manuals-reveal-terrorist-group-security-protocols/.

11 For helpful examples, see James P. Farwell, 'The Media Strategy of ISIS,' *Survival* 56, no. 6 (November 2, 2014): 49–55, doi:10.1080/00396338.2014.985436; Doina Chiacu, 'U.S. NSA Chief Says Monitoring Tech-Savvy Islamic State,' *Reuters*, September 16, 2014, http://www.reuters.com/article/us-cybersecurity-usa-islamic-state-idUSKBN0HB22A20140916; Brian Nussbaum, 'Thinking About ISIS And Its Cyber Capabilities: Somewhere Between Blue Skies and Falling Ones,' November 29, 2015, http://cyberlaw.stanford.edu/blog/2015/11/thinking-about-isis-and-its-cyber-capabilities-somewhere-between-blue-skies-and-falling; Benjamin Runkle, 'Is the Islamic State a Cyber Threat?,' *War on the Rocks*, September 9, 2015, http://warontherocks.com/2015/09/is-the-islamic-state-a-cyber-threat/; Michael Sheetz, 'How ISIS Is Using High-Tech Tools for Planning and Propaganda,' *The Fiscal Times*, December 4, 2015, http://www.thefiscaltimes.com/2015/12/04/How-ISIS-Using-High-Tech-Tools-Planning-and-Propaganda; Ashish Sen, 'How Do You Disrupt ISIS' Social Media Strategy and Safeguard Freedoms?,' *Atlantic Council*, January 21, 2016, http://www.atlanticcouncil.org/blogs/new-atlanticist/how-do-you-disrupt-isis-social-media-strategy-and-safeguard-freedoms; Manuel R. Torres-Soriano, 'The Caliphate Is Not a Tweet Away: The Social Media Experience of Al Qaeda in the Islamic Maghreb,' *Studies in Conflict & Terrorism* 0, no. ja (March 1, 2016): 1–35, doi:10.1080/1057610X.2016.1159430.

demonstrated in this paper, social media use, at least in the way that it is traditionally viewed, is a poor proxy for an actor's technical proficiency or cyber capabilities, and under certain conditions may actually highlight actor insecurity rather than competence.

Despite sharing some basic characteristics, social media activity does not translate frictionlessly into cyber capability. Each environment faces distinct problems and requires different tools and skills. A non-state or even a state actor's social media prowess is not a strong indicator of its technical proficiency or cyber capabilities. In fact, in many cases, social media use and its bidirectional nature can actually make a target *more* vulnerable. What is overlooked is that social media does play a role in cyber operations, just not the one that is often acknowledged. Social media's military utility extends far beyond broadcasting and counter-messaging operations. Social media operations can have value at the operational and tactical levels, and directly contribute to the effectiveness of Cyber Intelligence, Surveillance, and Reconnaissance (Cyber ISR) and Cyber Operational Preparation of the Environment (Cyber OPE).[12] For example, gathering direct content and metadata can reveal a target's specific software and hardware configuration or even its physical location. Social media can also provide a useful attack platform for the targeted delivery of a capability and an alternative command and control (C2) mechanism.[13] Thinking strategically about the use of social media in terms of active information-gathering, phishing, spamming, offensive cyber delivery methods, and targeted network degradation may provide a key advantage during conflict.

The structure of this paper is as follows. Section two outlines similarities and differences between social media and cyber security. Section three categorises different military social media operations into three types: information-gathering (IGMO), defensive operations (DeSMO), and offensive social media operations (OSMO). Section three also outlines key variables for social media platforms (e.g. type of content, filtering tools) and target actors (e.g group cohesion, size) to show that there is an important interaction between the type of social media operation, the type of platform, and the target actor's characteristics. Simply put, certain types of groups and social media platforms are more or less vulnerable to certain types of military social media operations. Finally, the paper ends by offering specific conclusions and recommendations for policymakers and academics.

2. SOCIAL MEDIA AND CYBER

Social media use is an increasingly important political and social science research area.[14] Domestically and internationally, social media and networked systems are being deployed to organise anti-government dissent, spread disaster information, enhance political campaigning,

12 'JP 3-12(R), Cyberspace Operations' (Department of Defense, February 5, 2013).

13 James C. Foster, 'The Rise Of Social Media Botnets,' *Dark Reading*, July 7, 2015, http://www.darkreading.com/attacks-breaches/the-rise-of-social-media-botnets/a/d-id/1321177; Spencer Ackerman, 'Pentagon Admits It Is 'Looking to Accelerate' Cyber-Attacks against Isis,' *The Guardian*, February 29, 2016, sec. World news, http://www.theguardian.com/world/2016/feb/29/pentagon-admits-cyber-attacks-against-isis; David E. Sanger and Nicole Perlroth, 'Iranian Hackers Attack State Dept. via Social Media Accounts,' *The New York Times*, November 24, 2015, http://www.nytimes.com/2015/11/25/world/middleeast/iran-hackers-cyberespionage-state-department-social-media.html.

14 Nils B. Weidmann, 'Communication, Technology, and Political Conflict Introduction to the Special Issue,' *Journal of Peace Research* 52, no. 3 (May 1, 2015): 263–68, doi:10.1177/0022343314559081.

and magnify the effects of terror recruitment campaigns.[15] While existing studies have closely examined social media use during periods of civil unrest and, more recently, in post-conflict reconstruction, its use for operational planning and specifically during conflicts is still under-theorised.[16] As demonstrated in Table 1, there are several recent examples of intrastate and interstate conflict where actors have deployed social media operations.[17] The selected cases are meant only to highlight useful examples and are not a representative sample of all potential cases. Two examples are worth discussing in greater detail.

TABLE 1: SOCIAL MEDIA USE

Interstate Conflict	Intrastate Conflict	Non-Conflict Operation Areas
Russia-Ukraine	ISIS Syria Libya Egypt Anonymous	Bin Laden raid

A. Existing social media use

1) Russia-Ukraine

Social media use in the Ukraine conflict demonstrates the increasing importance of states supplementing conventional capabilities with social media operations.[18] Social media platforms have been used by Russian military forces, intelligence agencies, and proxies to conduct information operations and for targeting and operational planning purposes.[19] Ukrainian military forces, proxies, and civilians have similarly deployed social media to spread information or gain an advantage.

15 See a good overview in Pablo Barberá and Thomas Zeitzoff, 'The New Public Address System: Why Do World Leaders Adopt Social Media?,' 2016, http://pablobarbera.com/static/world_leaders_paper.pdf; David C. Benson, 'Why the Internet Is Not Increasing Terrorism,' *Security Studies* 23, no. 2 (April 3, 2014): 293–328, doi:10.1080/09636412.2014.905353.

16 Thomas Zeitzoff, 'Using Social Media to Measure Conflict Dynamics: An Application to the 2008-2009 Gaza Conflict,' *Journal of Conflict Resolution*, June 20, 2011, 0022002711408014, doi:10.1177/0022002711408014; Jacob N. Shapiro and David A. Siegel, 'Coordination and Security How Mobile Communications Affect Insurgency,' *Journal of Peace Research* 52, no. 3 (May 1, 2015): 312–22, doi:10.1177/0022343314559624; Thomas Elkjer Nissen, *#TheWeaponizationOfSocialMedia: @ Characteristics_of_ Contemporary_Conflicts* (Royal Danish Defence College, 2015).

17 Doug Gross, 'Twitter User Unknowingly Reported Bin Laden Attack,' CNN, May 2, 2011, http://www.cnn.com/2011/TECH/social.media/05/02/osama.twitter.reports/index.html; 'Mapping the Syrian Conflict with Social Media,' Crisis.Net, 2014, http://crisis.net/projects/syria-tracker/; Masudul Biswas and Carrie Sipes, 'Social Media in Syria's Uprising and Post-Revolution Libya: An Analysis of Activists' and Blogger's Online Engagement,' Fall 2014, http://www.arabmediasociety.com/articles/downloads/20140925085334_BiswasSipes_SocialMedia_Final.pdf; Paul Roderick Gregory, 'Inside Putin's Campaign Of Social Media Trolling And Faked Ukrainian Crimes,' *Forbes*, May 11, 2014, http://www.forbes.com/sites/paulroderickgregory/2014/05/11/inside-putins-campaign-of-social-media-trolling-and-faked-ukrainian-crimes/; Dmitry Volchek and Claire Bigg, 'Ukrainian Bloggers Use Social Media to Track Russian Soldiers Fighting in East,' *The Guardian*, June 3, 2015, sec. World news, http://www.theguardian.com/world/2015/jun/03/bloggers-social-media-russian-soldiers-fighting-in-ukraine.

18 See above. Also see Kenneth Geers, *Cyber War in Perspective: Russian Aggression against Ukraine* (Tallinn: NATO CCD COE Publications, 2015); 'Twitter's Role in Modern Warfare,' *BBC News*, March 21, 2016, http://www.bbc.com/news/technology-35842265.

19 There are actor attribution issues but for the purposes of this paper it is sufficient to lump military forces together with intelligence forces, state-sponsored proxies, and activists. The key point is to highlight how social media is now a fundamental element of modern conflict zones.

A key part of Russia's strategy is to use social media platforms for military disinformation and propaganda campaigns.[20] The Russian government employs citizens as a 'troll army', consisting of social media users that inundate websites with pro-Putin rhetoric.[21] These trolls have been increasingly active in the lead up to major Russian foreign policy initiatives, including those in Crimea. Russian trolls can be savvy and technically sophisticated, and are capable of orchestrating advanced information campaigns while working from Russian territory.[22]

Russian military units have also been active in Ukraine, as evidenced by numerous incidents where Russian soldiers posted geotagged content (e.g., photos of weaponry) and commentary (referring to active fighting in Ukraine) to Instagram.[23] Through social media, reporters and academics have been able to document Russian military equipment deployed in places like Crimea and Ukraine.[24] Ukrainian civilians have also used social media to effectively communicate events as they are transpiring. For example, civilians have used social media to track Russian soldiers and to signal for help when caught between Ukrainian soldiers and pro-Russian separatists.[25]

2) ISIS

Social media provides ISIS with a flexible and streamlined set of tools for creating and distributing videos, images, and other content. ISIS routinely uses multiple social media platforms to broadcast anti-United States propaganda.[26] Inherent network effects then magnify the reach and effect of this propaganda. Social media also provides ISIS with a valuable means of engaging in targeted recruitment campaigns and attempts to radicalise target populations.[27]

20 Roman Skaskiw, 'Nine Lessons of Russian Propaganda | Small Wars Journal,' *Small Wars Journal*, March 27, 2016, http://smallwarsjournal.com/jrnl/art/nine-lessons-of-russian-propaganda.

21 Daisy Sindelar, 'The Kremlin's Troll Army,' *The Atlantic*, August 12, 2014, http://www.theatlantic.com/international/archive/2014/08/the-kremlins-troll-army/375932/.

22 Adrian Chen, 'The Agency,' *The New York Times*, June 2, 2015, http://www.nytimes.com/2015/06/07/magazine/the-agency.html.

23 Oscar Jonsson and Robert Seely, 'Russian Full-Spectrum Conflict: An Appraisal After Ukraine,' *The Journal of Slavic Military Studies* 28, no. 1 (January 2, 2015): 1–22, doi:10.1080/13518046.2015.998118 ; Max Seddon, 'Does This Soldier's Instagram Account Prove Russia Is Covertly Operating In Ukraine?,' *BuzzFeed*, July 30, 2014, http://www.buzzfeed.com/maxseddon/does-this-soldiers-instagram-account-prove-russia-is-covertl.

24 Jenny Hauser, 'Speed in Context: Real-Time News Reporting and Social Media,' 2014, http://arrow.dit.ie/aaschmedcon/36/; Maksymilian Czuperski et al., *Hiding in plain sight: Putin's war in Ukraine*, 2015, https://www.dropbox.com/s/7uzlm8aspdls5wh/Hiding-in-Plain_Sight_0527.pdf?raw=1.

25 Jenny Hauser, 'Speed in Context: Real-Time News Reporting and Social Media,' 2014, http://arrow.dit.ie/aaschmedcon/36/; Maksymilian Czuperski et al., *Hiding in plain sight: Putin's war in Ukraine*, 2015, https://www.dropbox.com/s/7uzlm8aspdls5wh/Hiding-in-Plain_Sight_0527.pdf?raw=1.

26 P.W. Singer and Emerson Brooking, 'Terror on Twitter,' *Popular Science*, December 11, 2015, http://www.popsci.com/terror-on-twitter-how-isis-is-taking-war-to-social-media; Brendan I. Koerner, 'Why ISIS Is Winning the Social Media War—And How to Fight Back,' WIRED, March 29, 2016, http://www.wired.com/2016/03/isis-winning-social-media-war-heres-beat/; J.M. Berger, 'How Terrorists Recruit Online (and How to Stop It),' *The Brookings Institution*, November 9, 2015, http://www.brookings.edu/blogs/markaz/posts/2015/11/09-countering-violent-extremism-online-berger; Daveed Gartenstein-Ross, 'The Social Science of Online Radicalization,' *War on the Rocks*, October 29, 2015, http://warontherocks.com/2015/10/the-social-science-of-online-radicalization/.; Klint Finley, 'It'd Be Great to Kick ISIS Offline—If It Were Possible,' *WIRED*, March 30, 2016, http://www.wired.com/2016/03/how-is-isis-online/.

27 Jytte Klausen, 'Tweeting the Jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq,' *Studies in Conflict & Terrorism* 38, no. 1 (January 2, 2015): 1–22, doi:10.1080/1057610X.2014.974948.; Haroro J. Ingram, 'Three Traits of the Islamic State's Information Warfare,' *The RUSI Journal* 159, no. 6 (November 2, 2014): 4–11, doi:10.1080/03071847.2014.990810.; Sabrine Saad Stéphane Bazan, 'Infowar on the Web: When the Caliphate Goes Online,' 2015, doi:10.13140/RG.2.1.1851.5043.

However, it is important to note that, like the social media activity in Ukraine, ISIS's social media activity is bidirectional. All parties can use social media for information-gathering and targeting purposes. For example, the United States Air Force used social media data posted by an ISIS supporter to target an ISIS military compound.[28] More recently, the United States Department of Defense has been engaged in ongoing social media and cyber operations against online ISIS targets.[29] Open source investigators have also successfully mapped the Twitter network of known ISIS supporters by analysing commonly used location and content data.[30] Even other non-state actors have similarly used social media to target and report ISIS social media accounts and websites.[31]

B. Social media meets cyber operations

The high profile nature and rise of social media activity by states and especially non-state actors has recently drawn the attention of those interested in cyber security. Specifically, commentators and researchers appear to view social media's relationship to cyber operations primarily in one of two ways. First, some observers have stretched the concept of cyber operations or cyber power to explicitly include social media activity.[32] Under this view, social media prowess becomes a primary example of an actor engaging in cyber operations.[33] Cyber technology and cyber operations then include a variety of different operations such as viral messaging on social media platforms, building internal messaging apps, intragroup operational security, deploying Distributed Denial of Service (DDoS) capabilities or even the deployment and use of advanced offensive cyber capabilities to achieve physical effects.

The second view maintains a narrower concept of cyber operations but still views social media activity or prowess as having a positive relationship with cyber capabilities.[34] Under this second view, social media operations are not synonymous with cyber operations but are instead an indicator of an actor's cyber capabilities. Actors that are successful at engaging in social

28 Walbert Castillo, 'U.S. Bombs ISIS Using Social Media Intel,' *CNN*, June 5, 2015, http://www.cnn.com/2015/06/05/politics/air-force-isis-moron-twitter/index.html.

29 Ackerman, 'Pentagon Admits It Is 'Looking to Accelerate' Cyber-Attacks against Isis'; Koerner, 'Why ISIS Is Winning the Social Media War—And How to Fight Back'; Christina Nemr, 'Strategies to Counter Terrorist Narratives Are More Confused than Ever,' *War on the Rocks*, March 15, 2016, http://warontherocks.com/2016/03/strategies-to-counter-terrorist-narratives-are-more-confused-than-ever/; Jared Cohen, 'Digital Counterinsurgency,' *Foreign Affairs*, December 2015, https://www.foreignaffairs.com/articles/middle-east/digital-counterinsurgency; Kimberly Dozier, 'Anti-ISIS-Propaganda Czar's Ninja War Plan: We Were Never Here.,' *The Daily Beast*, March 15, 2016, http://www.thedailybeast.com/articles/2016/03/15/obama-s-new-anti-isis-czar-wants-to-use-algorithms-to-target-jihadis.html.

30 JM Berger and Jonathan Morgan, 'The ISIS Twitter Census Defining and Describing the Population of ISIS Supporters on Twitter' (Washington, D.C: Brookings, March 2015), http://www.brookings.edu/~/media/research/files/papers/2015/03/isis-twitter-census-berger-morgan/isis_twitter_census_berger_morgan.pdf.

31 David Auerbach, 'The Hacktivist War on ISIS?,' *Slate*, December 10, 2015, http://www.slate.com/articles/technology/bitwise/2015/12/ghostsecgroup_is_taking_on_isis_it_s_not_clear_they_re_helping.html.

32 See footnote 12; Doina Chiacu, 'U.S. NSA Chief Says Monitoring Tech-Savvy Islamic State'; Brian Nussbaum, 'Thinking About ISIS And Its Cyber Capabilities'; Michael Sheetz, 'How ISIS Is Using High-Tech Tools for Planning and Propaganda'; Sen, 'How Do You Disrupt ISIS' Social Media Strategy and Safeguard Freedoms?'; Torres-Soriano, 'The Caliphate Is Not a Tweet Away'; Warwick Ashford, 'Social Media the Main Cyber Terror Threat Facing the UK, Says Former MI6 Officer,' *Computer Weekly*, October 16, 2015, http://www.computerweekly.com/news/4500255638/Social-media-the-main-cyber-terror-threat-facing-the-UK-says-former-MI6-officer.

33 Here social media 'use' and 'prowess' are largely used interchangeably. It is not always clear whether the people who make this first type of link between social media and cyber capabilities are addressing *any* use of social media or just instances of highly effective use.

34 Michael Sheetz, 'How ISIS Is Using High-Tech Tools for Planning and Propaganda'; Sen, 'How Do You Disrupt ISIS' Social Media Strategy and Safeguard Freedoms?'

media operations are also viewed to be broadly technically competent even to the degree of engaging in cyber operations. This argument has recently been used in debates surrounding the capabilities of ISIS, Anonymous, and Iranian forces.[35]

Both avenues of argument must logically rely on at least an implicit assumption that the same skills that allow actors to be successful at social media operations also enable them to be successful at other technical skills or even offensive and defensive cyber operations. In the first argument, social media skills and cyber security skills match one-to-one. Broadening the concept of cyber capabilities to include social media operations means that, by definition, the actor that just engaged in successful social media operations is now 'cyber capable.' Unfortunately, this conceptual stretching is not only tautological but is also not particularly helpful. At best, it indicates that the actor is capable of deploying only one minor type of cyber operations, social media operations. The argument is agnostic on the real question of whether that actor is able to successfully engage in defensive and offensive military cyber missions. At worst, this first type of argument stretches the concept of cyber capability to the point of incoherence.

The second avenue of argument initially appears more promising. Perhaps there are shared traits or skillsets between successful social media operations and cyber capabilities. If so, then successful social media operations may be a useful proxy variable for an actor's latent cyber capability. Even a weak positive relationship may demonstrate that an actor that engages in successful social media operations is more likely than other actors to have functioning cyber capabilities. Unfortunately, the link between social media and cyber in terms of shared traits and skills remains to be demonstrated.[36]

At the most basic level, both social media operations and cyber operations share common elements. First, they both rely heavily on building up skilled human capital. Second, they both involve some degree of technical or computer knowledge. Third, they both involve some knowledge of network effects. Fourth, they both involve elements of working in real time. Fifth, they both involve working within limitations set by a system. In the case of social media operations, these limitations are set by the specific platform being used. In the case of cyber operations, the limitations are primarily dictated by the target's systems and the nature of the specific vulnerability that is being exploited. However, even at this most basic level the differences in terms of scale and degree of skill, technical knowledge, network effects, and system limitations are extremely large.

The technical knowledge involved in social media operations is primarily focused on deploying an already publically or commercially developed tool. The actor only needs to understand how

35 See footnote 12; Ibid.; Doina Chiacu, 'U.S. NSA Chief Says Monitoring Tech-Savvy Islamic State'; Brian Nussbaum, 'Thinking About ISIS And Its Cyber Capabilities'; Michael Sheetz, 'How ISIS Is Using High-Tech Tools for Planning and Propaganda'; Sen, 'How Do You Disrupt ISIS' Social Media Strategy and Safeguard Freedoms?'; Torres-Soriano, 'The Caliphate Is Not a Tweet Away'; Warwick Ashford, 'Social Media the Main Cyber Terror Threat Facing the UK, Says Former MI6 Officer'; Meg King and Grayson Clary, 'Opinion: The Shocking Mediocrity of Islamic State 'Hacker' Junaid Hussain,' *Christian Science Monitor*, October 26, 2015, http://www.csmonitor.com/World/Passcode/Passcode-Voices/2015/1026/Opinion-The-shocking-mediocrity-of-Islamic-State-hacker-Junaid-Hussain; 'Twitter's Role in Modern Warfare'; Elias Groll, 'Welcome to the Future of War: ISIS Has a Smartphone App,' *Foreign Policy*, December 8, 2015, https://foreignpolicy.com/2015/12/08/welcome-to-the-future-of-war-isis-has-a-smartphone-app/; 'Who's at the Controls of Iran's Bot Army?,' *BBC News*, March 16, 2016, http://www.bbc.com/news/blogs-trending-35778645.

36 For space reasons, this paper will only briefly cover a few key similarities and differences between social media operations and cyber operations.

to deploy the tool, but does not need to have working knowledge of how that tool was built or how it functions. The same argument applies to the degree of skill that is needed and knowledge of systems and networks effects. Moreover, unlike social media, cyber operations involve a strategic interaction between attackers and defenders.[37] Defenders are able to react and respond in a way that requires a high degree of skill and time to successfully overcome.

Even bracketing skill comparisons, the two types of operations involve antithetical problems. In almost all cases, social media platforms ensure access by default. An actor has direct access to a target or a specific network because it is a built-in property of the platform. For example, social media platforms such as Twitter or Facebook are public by default. In the case of cyber operations, the key problem is to overcome restricted access. The target in a cyber operation is restricting access by default whereas with social media the target welcomes the actor. Similarly, in the social media case the actor wants to magnify and broadcast a message or type of content using network properties. In the offensive cyber case, the actor often wants to conceal and narrow the scope of the operation.

There is a relationship between social media and cyber operations, just not the one that is traditionally acknowledged. Social media operations directly contribute to Cyber Intelligence, Surveillance, and Reconnaissance (Cyber ISR) and Cyber Operational Preparation of the Environment (Cyber OPE). As will be demonstrated below, social media operations can be valuable at the operational and tactical levels. Operations can reveal useful information for weaponeering a specific cyber capability against a specific target.[38] Social media operations may also reveal both a means of capability deployment against a target's systems and alternative mechanisms for command and control. Actors that are highly active on social media may actually be increasing their vulnerability to offensive cyber capabilities by revealing target-specific information and widening the attack surface.

3. A FRAMEWORK FOR SOCIAL MEDIA OPERATIONS

The previous section demonstrated a number of ways that social media has been used in existing conflict zones and hinted at social media's usefulness as a complement to an actor's existing cyber capabilities. This section further unpacks social media operations (SMO) into its component types and directly links each to cyber operations. Social media operations consist of three distinct types: information-gathering, defence, and offense.

A. Information-gathering media operations

Information-gathering media operations (IGMO) focus on passive information-gathering. As demonstrated in the Ukraine, ISIS, and Bin Laden Raid cases, passive information-gathering

37 Drew Herrick and Trey Herr, 'Combating Complexity.'
38 See also, Ibid.

can be used for monitoring adversary activities and for targeting.[39] Through IGMO, military and intelligence forces are not interacting with known social media actors but instead are passively monitoring and documenting social media activity. IGMO focuses on two types of data: (1) direct data collection (the content displayed on social media); and (2) metadata collection (technical details related to the characteristics of social media users and the mechanics of their social media use). Direct data collection allows access to the actual content displayed on social media services.

Metadata collection is not as qualitatively rich as direct data collection, but can reveal important details regarding a population or target's location, the time of day that the target is active, the target's social graph (network connections), specific applications that the target is using to access services, whether the target is using a mobile device, and in some cases even the specific hardware and software configuration of the device that the target is using.[40]

This information can then be refined for non-kinetic purposes such as cyber ISR or OPE, or for kinetic targeting (e.g., physical destruction). Whether used for direct data or metadata collection, IGMO can be a useful complement to other information collection activities. The primary risk to using IGMO for these purposes is that strategic and competent adversaries may intentionally cleanse or manipulate social media information in order to mislead those trying to monitor various sources. For example, strategic actors that realise they are being observed may take steps to mask their location, use automation to schedule activity, or intentionally communicate false information to influence the observer forces to act in a certain way.[41] Similarly, maintaining the ability to engage in IGMO requires that an adversary's social network accounts be left up and running.[42] Legal attempts to cut off an adversary from using social media platforms directly trades off with the ability to gather key information.

39 Jamie Bartlett and Louis Reynolds, *The State of the Art 2015: A Literature Review of Social Media Intelligence Capabilities for Counter-Terrorism* (London: Demos, 2015), http://www.demos.co.uk/wp-content/uploads/2015/09/State_of_the_Arts_2015.pdf; Aliya Sternstein, 'Pentagon Mapmakers Are Using Social Media to Chart Syrians' Exodus,' *Defense One*, March 20, 2016, http://www.defenseone.com/technology/2016/03/pentagons-cartographers-are-mapping-syrias-exodus-thanks-social-media/126808/; Patrick M. Gillen, 'Real-Time Detection of Operational Military Information in Social Media' (Thesis, Monterey, California: Naval Postgraduate School, 2015), null/handle/10945/47261; Swati Agarwal, Ashish Sureka, and Vikram Goyal, 'Open Source Social Media Analytics for Intelligence and Security Informatics Applications,' in *Big Data Analytic*s, ed. Naveen Kumar and Vasudha Bhatnagar, Lecture Notes in Computer Science 9498 (Springer International Publishing, 2015), 21–37, http://link.springer.com/chapter/10.1007/978-3-319-27057-9_2; Robert Chesney, 'Anonymous vs ISIS Online: Pondering the Intelligence Impact of Social Media Takedowns,' *Lawfare*, November 18, 2015, https://www.lawfareblog.com/anonymous-vs-isis-online-pondering-intelligence-impact-social-media-takedowns; Alastair Paterson, 'Using an Attacker's 'Shadow' to Your Advantage | SecurityWeek.Com,' *Security Week*, November 5, 2015, http://www.securityweek.com/using-attackers-shadow-your-advantage.

40 Bo Zhao and Daniel Sui, 'True Lies in Big Data: Detecting Location Spoofing in Social Media,' *Journal of Spatial Information Science*, 2016, http://www.josis.org/index.php/josis/article/viewArticle/273.

41 Michela Del Vicario et al., 'The Spreading of Misinformation Online,' *Proceedings of the National Academy of Sciences* 113, no. 3 (January 19, 2016): 554–59, doi:10.1073/pnas.1517441113.

42 Patrick Tucker, 'Twitter Steps Up Efforts To Combat ISIS,' *Defense One*, February 5, 2016, http://www.defenseone.com/technology/2016/02/twitter-steps-efforts-combat-isis/125739/; J.M. Berger and Heather Perez, 'The Islamic State's Diminishing Returns on Twitter: How Suspensions Are Limiting the Social Networks of English-Speaking ISIS Supporters' (Washington, D.C: George Washington University, February 2016), https://cchs.gwu.edu/sites/cchs.gwu.edu/files/downloads/Berger_Occasional%20Paper.pdf.

B. Defensive social media operations

Defensive social media operations (DeSMO) involve using social media in a more active way than IGMO, but not as active as OSMO. Actors can use social media as a broadcasting platform to conduct counter-messaging or counter-propaganda activities.[43] As demonstrated in the Russian troll and ISIS cases, social media can be used effectively to widely broadcast information to otherwise difficult-to-reach audiences. In fact, US government agencies are already using social media services to counteract known propaganda and radicalisation campaigns.[44] However, existing operations are extremely limited and, at best, produce minor effects.[45]

Despite its value, DeSMO has the potential downside of providing an adversary with direct data collection opportunities and metadata that would otherwise not be revealed. Put differently, engaging in DeSMO activities allows the adversary to engage in IGMO or even OSMO. While this information can be shielded, its emission is nonetheless a risk that must be acknowledged. By engaging in counter-messaging, the actors involved are revealing information about, for example, their own capabilities, location, or system configurations. DeSMO does not play a direct role in terms of cyber operations, but has been acknowledged as a key component of de-radicalisation campaigns.

C. Offensive social media operations

Social media operations are commonly viewed as a broadcasting or counter-narrative tool; DeSMO under this paper's new framework. More recently, social media operations as a passive information-gathering (or IGMO) tool have received some attention as the conversation surrounding ISIS and online radicalisation has subtly shifted from 'shut it down' towards a monitoring mentality.[46] Instead of actively shuttering known ISIS accounts and websites, intelligence agencies and even non-governmental actors can passively observe and analyse their content.

43 David P. Fidler, 'Countering Islamic State Exploitation of the Internet' (Washington, D.C: Council on Foreign Relations, June 2015), http://www.cfr.org/cybersecurity/countering-islamic-state-exploitation-internet/p36644; Dann Albright, 'How Social Media Is The Newest Military Battleground,' *MakeUseOf*, February 19, 2015, http://www.makeuseof.com/tag/social-media-newest-military-battleground/; P.W. Singer and Emerson Brooking, 'Terror on Twitter'; David Ensor, 'How Washington Can Win the Information War,' *Foreign Policy*, December 14, 2015, https://foreignpolicy.com/2015/12/14/how-washington-can-win-the-information-war/.

44 Greg Miller and Karen DeYoung, 'Obama Administration Plans Shake-up in Propaganda War against ISIS,' *The Washington Post*, January 8, 2016, https://www.washingtonpost.com/world/national-security/obama-administration-plans-shake-up-in-propaganda-war-against-the-islamic-state/2016/01/08/d482255c-b585-11e5-a842-0feb51d1d124_story.html.

45 Koerner, 'Why ISIS Is Winning the Social Media War—And How to Fight Back'; Charlie Winter and Jordan Bach-Lombardo, 'Why ISIS Propaganda Works,' *The Atlantic*, February 13, 2016, http://www.theatlantic.com/international/archive/2016/02/isis-propaganda-war/462702/; Patrick Tucker, 'Pentagon: State Doesn't Have Enough People Tweeting At ISIS,' *Defense One*, October 22, 2015, http://www.defenseone.com/technology/2015/10/pentagon-state-doesnt-have-enough-people-tweeting-isis/123063/; Christina Nemr, 'Strategies to Counter Terrorist Narratives Are More Confused than Ever'; Jared Cohen, 'Digital Counterinsurgency.'

46 Julia Greenberg, 'Facebook And Twitter Face Tough Choices As ISIS Exploits Social Media to Spread Its Message,' *WIRED*, November 21, 2015, http://www.wired.com/2015/11/facebook-and-twitter-face-tough-choices-as-isis-exploits-social-media/; Albanesius, 'Obama: Intelligence Officials 'Constantly' Monitor Social Media Posts,' *PCMAG*, December 18, 2015, http://www.pcmag.com/article2/0,2817,2496846,00.asp.

Offensive social media operations (OSMO) are still largely ignored in existing research and in the cases discussed above.[47] OSMO includes activities conducted on social media platforms to actively gather information, conduct information campaigns, deliver precision cyber effects, and counter, degrade, deny, or destroy an adversary's social media capability. In these respects, social media's bidirectional nature can be used as a vector to target and attack adversaries by and through their own social media activity.

For example, OSMO can enable military forces or intelligence agencies to spam known actors or networks to increase the overall signal-to-noise ratio within a given social media environment. Depending on the specific filtering tools of the social media services being employed, strategic spamming may allow friendly forces to disrupt an adversary's social media use while still leaving the broader service and supporting networks up and running. Using OSMO in this manner would allow IGMO and DeSMO efforts to continue uninterrupted while still allowing for the disruption of an adversary's social media use. In cases where an adversary is using several social media platforms simultaneously, military forces can also selectively disrupt one platform to shift activity to another where they may have a larger comparative advantage.

Militaries can also use 'trolling' techniques to target normally unresponsive or inactive accounts. This more active form of engagement with an account may incentivise the target actor to lash out in response, thereby revealing more direct and indirect information. Similarly, social media can be used for phishing purposes. These specific techniques may especially benefit from deploying proxies or 'cyber mercenaries.'

Finally, social media platforms can be used as both an attack avenue for offensive cyber capabilities and as an alternative means for command and control (C2).[48] Since access is often built in by default, using social media as a delivery platform may reduce the cost and time associated with traditional ways of deploying offensive cyber capabilities.

D. Limitations and opportunities

Despite IGMO, DeSMO and OSMO yielding potentially valuable advantages, these benefits are not universal. First, social media operations only yield a benefit in conflict areas that already have a high degree of connectedness and social media activity. Trying to use social media techniques in non-networked environments will not be particularly fruitful. Second, social media operations are bidirectional; actively using social media might provide unintended benefits to an adversary. Third, social media operations are likely to involve very large networks, requiring a high degree of competency and sophistication to effectively monitor and influence. Finally, many social media operations will have to be conducted in real-time or near real-time, and effective operations will require continuous monitoring and response.

There are also non-network considerations that may limit the utility of social media operations. First, there is an intrinsic authenticity problem. Depending on whether the target is aware that they are under surveillance and the sophistication of their understanding of the social media environment, there may be significant uncertainty concerning the veracity of information

47 For a few examples, see Adam Weinstein, 'Here's How the US Should Fight ISIS With Social Media,' *WIRED*, March 12, 2015, http://www.wired.com/2015/03/heres-us-fight-isis-social-media/; Nissen, *#TheWeaponizationOfSocialMedia*; Heather M. Roff et al., 'Fight ISIS by Thinking Inside the Bot,' Slate, October 21, 2015, http://www.slate.com/articles/technology/future_tense/2015/10/using_chatbots_to_distract_isis_recruiters_on_social_media.html.

48 James C. Foster, 'The Rise Of Social Media Botnets.'

gathered. One other interesting point is the effect of group size. There may be good reasons to anticipate that faking information will be more difficult as social media groups grow over time.

Second, the effectiveness of social media operations is contingent on the characteristics of the specific social media service and the target group. There is a key interaction effect between the type of action chosen (IGMO, DeSMO, or OSMO), the specific attributes of the social media platform in use, and the specific dynamics of the target. This interaction effect then directly impacts the effectiveness (high, medium, or low) of a given option. For example, social media services with high entry costs and very good filtering tools will be highly resistant to network or even individual node spamming. Likewise, groups that have a high degree of familiarity with technology and have been using a specific social media service for a long time will be more resistant to certain types of operations. However, targets with low group cohesion, high turnover, and low familiarity with a given platform may be especially vulnerable to targeted social media operations. Low cohesion and high turnover mean that it is less likely that every actor within the group knows every other actor. Impersonation tactics may be particularly effective. Finally, there are significant regulatory, doctrinal, and structural issues that must be resolved if social media operations are going to be conducted by military forces or even intelligence agencies. Overall, these limitations restrict the use social media operations but do not eliminate their utility.

4. CONCLUSION

Social media use, as it is traditionally viewed, is a poor indicator of an actor's *true* technical ability, cyber capabilities, or 'cyber power.' Viewing social media operations as either a direct example of an actor's cyber operations in action or as a reliable proxy for latent cyber capabilities is misguided. Both options hinge on false assumptions about the relationship between social media and cyber operations. This paper has made two arguments. First, that if social media operations are to be directly connected to cyber operations then it is better to view those operations as complementary to an already existing cyber capability. Second, it has outlined a preliminary framework for social media operations that can be unpacked into three distinct types: information-gathering, defence, and offence. In short, social media operations provide potentially useful information for targeting purposes and defensive threat intelligence, and expand the attack surface.

Policymakers and academics should focus on the broader utility of social media operations for military effectiveness. How can social media operations be successfully integrated with existing cyber and information operations? Should states push for international norms or treaties that apply to the use of social media during peace and conflict? Can offensive strategies be developed to successfully counter social media use by an adversary?

Overall, successful social media operations may act as a powerful force multiplier for both conventional and cyber capabilities. Thinking seriously about the nature of social media operations may help inform the future direction of military force structure and policies surrounding how to counter violent state and non-state actors.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Influence Cyber Operations: The Use of Cyberattacks in Support of Influence Operations

Pascal Brangetto
NATO Cooperative Cyber Defence Centre of Excellence
Tallinn, Estonia
pascal.brangetto@ccdcoe.org

Matthijs A. Veenendaal
NATO Cooperative Cyber Defence Centre of Excellence
Tallinn, Estonia
matthijs.veenendaal@ccdcoe.org

Abstract: Information Warfare and Influence Operations are, in principle, intended to get your own message across or to prevent your adversary from doing so. However, it is not just about developing a coherent and convincing storyline as it also involves confusing, distracting, dividing, and demoralising the adversary. From that perspective, cyberspace seems to be ideal for conducting such operations that will have disruptive, rather than destructive outcomes.

The means through which influence can be exerted relies mostly on spreading information. However, there are more intrusive ways to influence specific audiences that remain in the information realm but are designed to change, compromise, inject, destroy, or steal information by accessing information systems and networks. This paper aims to tackle the following questions: when does influencing the behaviour of an audience become the primary effect of a cyber operation, and which cyber operations might qualify as such? We introduce the term Influence Cyber Operations (ICOs) to describe these actions in cyberspace.

In order to address these questions, and drawing from existing literature, this paper defines ICOs as a specific subset of Influence Operations. ICOs encompass activities undertaken in cyberspace affecting the logical layer of cyberspace with the intention of influencing attitudes, behaviours, or decisions of target audiences. To illustrate the case for ICOs, we comment on a broad range of techniques that can be used in order to conduct these, and discuss the accompanying policy frameworks.

Keywords: *influence operations, information operations, information warfare, strategic communications*

1. INTRODUCTION

Nations have always used information to enhance their goals and policies as conflicts have never been limited to the military realm.[1] Today, with its rapid expansion, cyberspace seems to be ideal for conducting Influence Operations, maybe even more than for conducting destructive operations.[2] As Tim Stevens puts it, 'cyber warfare of the future may be less about hacking electrical power grids and more about hacking minds by shaping the environment in which political debate takes place'.[3]

The objective of Influence Operations is predominantly to exert power by influencing the behaviour of a target audience; the ability for 'A to have B doing, to the extent that he can get B to do something that B would not otherwise do'.[4] Influence Operations are thus assumed to modify attitudes and shape opinions through the dissemination of information and conveying of messages.[5] However, there are more intrusive ways to influence a specific audience that remain in the information realm but can no longer be regarded as the application of soft power as they are no longer designed to achieve their objective solely through 'attraction'.[6] Cyberspace offers numerous possibilities for these kinds of coercive operations, which are designed to influence a target audience by changing, compromising, destroying, or stealing information by accessing information systems and networks.

The question then arises: when does influencing the behaviour of an audience become the primary effect of a cyber operation and which cyber operations might qualify as such? This paper addresses this question by describing the cyber aspects of Influence Operations and how their technical features may play an active role regardless of their content. We will therefore focus on the relevance of intrusive cyber operations to Influence Operations, for which we propose the term Influence Cyber Operations (ICO).

In this paper, the authors argue that coercive ICOs will become more prevalent because they offer the opportunity to undermine an opponent's credibility with little risk of escalation. When defining ICOs, we highlight the confusion pertaining to the terminology regarding Influence Operations (Section 2). The main attraction for the use of ICOs lies in the fact that they are generally limited in scope and difficult to attribute, thereby limiting the risks of escalation and countermeasures. This is especially reflected in the Russian approach to Information Warfare, which considers it as an instrument of hard power. By contrast, because of the importance

1 'The expansion of the domain of warfare is a necessary consequence of the ever-expanding scope of human activity, and the two are intertwined.' in Qiao Liang and Wang Xiangsui, *Unrestricted Warfare*, PLA Literature and Arts Publishing House, 1999, p. 189.

2 'Rather than a 'Cyber Armageddon' scenario that debilitates the entire US infrastructure, we envision something different. We foresee an ongoing series of low-to-moderate level cyber attacks from a variety of sources over time, which will impose cumulative costs on US economic competitiveness and national security' Statement of James Clapper for the Record Worldwide Threat Assessment of the US Intelligence Community Senate Armed Services Committee, 26 February 2015, p.1 http://www.dni.gov/files/documents/Unclassified_2015_ATA_SFR_-_SASC_FINAL.pdf. (All Internet resources were accessed 4 March 2016).

3 Ben Quinn, 'Revealed: the MoD's secret cyberwarfare programme', *The Guardian*, 16 March 2014, http://www.theguardian.com/uk-news/2014/mar/16/mod-secret-cyberwarfare-programme.

4 Here, we use the definition provided by Robert Dahl in his seminal article, 'The concept of Power', *Behavioural Science*, 2:3, July 1957.

5 William Hutchinson, *Influence Operations: Action and Attitude*, 2010. http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1032&context=isw.

6 Joseph Nye, 'Soft Power, the means to succeed in world politics', *Public Affairs* 2004, p. x.

Western democracies attach to issues of legality and transparency, their options for using ICOs remain, in principle, limited. Looking at the different approaches (Section 3), this paper then describes what ICOs look like and how they may be applied (Section 4) and provides conclusions and basic recommendations (Section 5).

2. THE DEFINITION CONUNDRUM

In 2007 Martin C. Libicki noted 'that well over a decade after the topic of information warfare broke out into the open, its conceptual underpinnings remain weak and largely unsatisfactory, with fierce battles raging over neologisms and definitions'.[7] Almost a decade later, progress on this issue remains slow. There is still a lack of consensus when it comes to defining all the elements that make up the strategic application of power in the information domain. Regarding the use of terms like Information Warfare (IW), Psychological Operations (PSYOPS), Influence Operations (IO), Strategic Communications (STRATCOM), Computer Network Operations (CNO), and Military Deception (MILDEC), there is a lot of confusion as there are numerous conflicting definitions, and these terms are used in different contexts to describe different objectives and actions.[8] When trying to make sense of the information domain it is therefore necessary to clarify and define the terminology that is used in this paper. The authors of this article do not, however, seek to provide any definitive answers on this issue.

The main reason for us to provide specific definitions is that Influence Operations are not limited to military operations, but can be part of any kind of conflict, including, for example, in the diplomatic arena. They are therefore part of a larger effort by nations to exert power over adversaries.

In principle, Influence Operations offer the promise of victory through:

> 'the use of non-military [non-kinetic], means to erode the adversary's willpower, confuse and constrain his decision-making, and undermine his public support, so that victory can be attained without a shot being fired'.[9]

They include all the efforts undertaken by states or any other groups to influence the behaviour of a target audience, in peacetime or during an armed conflict. It is therefore the umbrella term for all operations in the information domain, including all soft power activities. Although Influence Operations are, in principle, non-violent, they can be part of military operations.

In addition, Influence Operations are not solely confined to the application of soft power. They can also include clandestine and intrusive activities undertaken as part of an armed conflict or

7 Martin Libicki, *Conquest in Cyberspace, National Security and Information Warfare*, 2007, Cambridge University Press, p. 17.

8 This confusion is underlined by the definition of Information Warfare (IW) provided by RAND in Eric V. Larson, Richard E. Darilek, Daniel Gibran, Brian Nichiporuk, Amy Richardson, Lowell H. Schwartz, Cathryn Quantic Thurston, *Foundations of Effective Influence Operations A Framework for Enhancing Army Capabilities*, Rand Corporation, 2009 p 2. 'Information Warfare is conflict or struggle between two or more groups in the information environment' which is such a blanket definition that, although technically correct, it borders on being useless.

9 Anne Applebaum, Edward Lucas, *Wordplay and War Games*, 19 June 2015, http://www.cepa.org/content/wordplay-and-war-games.

military operation. This is in line with the definition we use in this paper, which includes the possibility of the use of intrusive cyber capabilities:

> 'Influence operations are the coordinated, integrated, and synchronized application of national diplomatic, informational, military, economic, and other capabilities in peacetime, crisis, conflict, and post-conflict to foster attitudes, behaviours, or decisions by foreign target audiences that further [a nation's] interests and objectives'.[10]

For the much-used term 'Information Operations', we rely on the US DoD definition, which defines it as a military capability that is:

> '[t]he integrated employment, during military operations, of information-related capabilities in concert with other lines of operations to influence, disrupt, corrupt, or usurp the decision making of adversaries and potential adversaries while protecting [its] own'.[11]

In our approach, Information Operations are therefore a subset of Influence Operations limited to military operations.

If Influence Operations are also understood to include intrusive operations, it becomes necessary to separate the 'apples' of information content from the 'apple carts' of information systems.[12] This is in line with Russian thinking on Information Warfare, which traditionally makes the distinction between 'informational-technical' and 'informational-psychological' activities. The semantic or cognitive actions (apples) consist mainly of attacks of information on information (typically narrative vs narrative) that affects the semantic layer of cyberspace. In other words, these are the activities in cyberspace that aim to produce content to create a crafted informational environment. These content-oriented activities can be defined as Inform & Influence Operations (IIOs) that we define as follows:

> 'Inform & Influence Operations are efforts to inform, influence, or persuade selected audiences through actions, utterances, signals, or messages'.[15]

Strategic communications (STRATCOM) and propaganda activities fall under this category, as well as the deliberate dissemination of disinformation to confuse audiences.

The 'apple carts' of Influence Operations concern the technical actions that target the logical

10 Eric V. Larson, Richard E. Darilek, Daniel Gibran, Brian Nichiporuk, Amy Richardson, Lowell H. Schwartz, Cathryn Quantic Thurston, *Foundations of Effective Influence Operations A Framework for Enhancing Army Capabilities*, RAND Corporation, 2009 p. 2.

11 US Department of Defense, Directive 3600.01. May 2, 2013. p.12.

12 This comparison is taken from Christopher Paul, *Information Operations, Doctrine and Practice, a Reference Handbook*, Praeger Security International, 2008, p. 37.

13 Timothy Thomas, *Recasting the Red Star*, Foreign Military Studies Office, Fort Leavensworth 2011, p. 138, http://fmso.leavenworth.army.mil/documents/RecastingRedStar_2015.pdf.

14 See Martin Libicki, *Conquest in Cyberspace, National Security and Information Warfare*, Cambridge University Press, pp. 24-25.

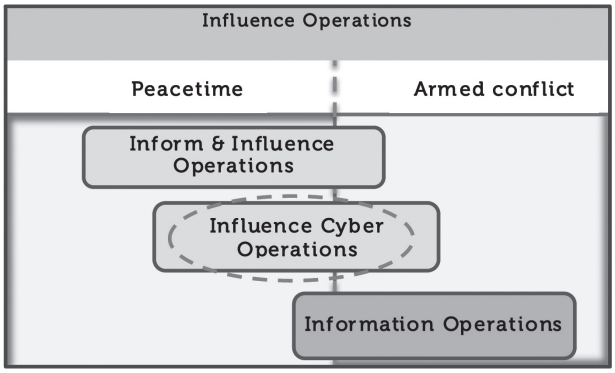
15 Isaac R. Porche III et.al, *Redefining Information Warfare Boundaries for an Army in a Wireless World*, Santa Monica 2013, Page xx.

layer of cyberspace and are designed to influence the behaviour of a target audience.[16] These actions are intrusive as they gain unauthorised access to networks and systems in order to destroy, change or add information. We use the term Influence Cyber Operations (ICOs) for these operations, which we define as follows:

> 'Operations which affect the logical layer of cyberspace with the intention of influencing attitudes, behaviours, or decisions of target audiences.'

ICOs are therefore activities undertaken in and through cyberspace and qualify as cyberattacks. For the purposes of this article, a cyberattack is understood to be '[a]n act or action initiated in cyberspace to cause harm by compromising communication, information or other electronic systems, or the information that is stored, processed or transmitted in these systems.'[17] By 'harm', in addition to physical damage, we also comprise the effects on information systems, hence 'direct or indirect harm to a communication and information system, such as compromising the confidentiality, integrity, or availability of the system and any information exchanged or stored.'[18]

FIGURE 1: INFLUENCE OPERATIONS SPECTRUM



3. POLICY FRAMEWORKS FOR INFLUENCE CYBER OPERATIONS

With the advent of the digital domain and the renewed interest in hybrid threats as a result of the Russian aggression against Ukraine in 2014 and its intervention in Syria in 2015, Influence Operations have received greater attention.[19] Influence Operations are an integral part of

16 We define the logical layer based on the definition of the syntactic layer provided by Martin Libicki as 'the various instructions and services that tell information systems what to do with information. [It] may be said to include operating systems (Oss) and applications. Network syntax clearly includes routing, but also access controls and security, directories, utility servers, and commonly used databases.' in Martin Libicki, *Supra*, p.25.

17 *NATO Report on Cyber Defence Taxonomy and Definitions*, Enclosure 1 to 6200/TSC FCX 0010/TT-10589/Ser: NU 0289.

18 *Ibid*.

19 Jan Joel Andersson and Thierry Tardy, 'Hybrid, what's in a name?', European Union Institute for Security Studies Brief 32, October 2015, http://www.iss.europa.eu/uploads/media/Brief_32_Hybrid_warfare.pdf; Sam Jones, 'Russia steps up Syria Cyber Assault', *Financial Times*, 19 February 2016, http://www.ft.com/intl/cms/s/0/1e97a43e-d726-11e5-829b-8564e7528e54.html#axzz41f9B2xIw.

hybrid warfare, which is the coordinated, overt, and covert use of a broad range of instruments, military and civilian, conventional and unconventional, in an ambiguous attack on another state.[20] Hybrid warfare provides many opportunities for the use of cyber capabilities as one of the broad range of possible non-kinetic or non-violent options. If the main goal of (political) Influence Operations outside of an armed conflict is to destabilise and confuse adversaries, then it could be effective to attack the opponent's digital infrastructure to undermine trust by compromising, altering, disrupting the digital services of both government and private sector through the use of malware.[21]

The strategic outlooks of nations on Influence Operations differ greatly. Where Russia and China have developed more integrated and holistic views, Western states, in general, tend to adopt a much more compartmentalised approach. Given these profound differences in the approaches of Russia and most NATO-members, we will analyse the contradicting strategies and ways in which Influence Operations are conducted.

A. The Russian approach

Russia, more than any other actor, seems to have devised a way to integrate cyber operations into a strategy capable of achieving political objectives.[22] Russia's approach in its power struggle with NATO and the West is based on the acknowledgement that it cannot match the military power of NATO.[23] Strategic advantages must therefore be achieved without provoking an armed response from the Alliance. This is a core element of Russian security policy which is based on the assumption that conflicts between developed nations must remain below the threshold of an armed conflict, or at least below the threshold where it is actually proclaimed to be an armed conflict. This strategy is exemplified by the Gerasimov doctrine (Russian non-linear war[24]) which posits that '[t]he role of non-military means of achieving political and strategic goals has grown, and, in many cases, they have exceeded the power of weapons in their effectiveness'.[25] Hence, a greater reliance on the information domain is obvious.

In the Russian view, Information Warfare is conducted in *peacetime*, in the *prelude to war* and in *wartime* in a coherent manner.[26] Information warfare uses:

20 NATO is also addressing this challenge so that it is 'able to effectively address the specific challenges posed by hybrid warfare threats, where a wide range of overt and covert military, paramilitary, and civilian measures are employed in a highly integrated design.', NATO Wales Summit Declaration, 5 September 2014.

21 Roland Heickerö, *Emerging Cyber Threats and Russian Views on Information Warfare and Information Operations*, FOI-R-2970-SE, p. 20.

22 James J. Wirtz, 'Cyber War and Strategic Culture: The Russian Integration of Cyber Power into Grand Strategy' in Kenneth Geers (Ed.), *Cyber War in Perspective: Russian Aggression against Ukraine* p. 21.

23 Keir Giles, *Russia's New' Tools for Confronting the West Continuity and Innovation in Moscow's Exercise of Power*, Chatham House, March 2016, p. 26.

24 Peter Pomerantsev, 'How Putin is reinventing warfare', *Foreign Policy*, 5 May 2014, http://foreignpolicy.com/2014/05/05/how-putin-is-reinventing-warfare/.

25 See Mark Galeotti' blog, *In Moscow Shadows*, https://inmoscowsshadows.wordpress.com/2014/07/06/the-gerasimov-doctrine-and-russian-non-linear-war/.

26 A.N. Limno, M.F. Krysanov, 'Information Warfare and Camouflage, Concealment and Deception', *Military Thought*, 2003, vol. 12, no. 2. 'Russian […] writing on the subject has more explicitly retained the more holistic and integrated view of information warfare'. Keir Giles and William Haggestad II, 'Divided by a Common Language: Cyber Definitions in Chinese, Russian and English', in Karlis Podins, Jan Stinissen, Markus Maybaum (Eds.), *5th International Conference on Cyber Conflict Proceedings*, 2013, NATO CCDCOE Publications, Tallinn, p. 422.

> ‘all the means and methods of impacting information, information-psychological, and information-technological objects and information resources to achieve the objectives of the attacking side’.[27]

These include intelligence, counterintelligence, deceit, disinformation, electronic warfare, debilitation of communications, degradation of navigation support, psychological pressure, degradation of information systems, and propaganda. In this context, be it distributed denial of service attacks (DDoS), advanced exploitation techniques, or RT television, all contribute to the same goals.[28]

From this perspective, using intrusive ICOs as part of a broader Influence Operations strategy makes perfect sense. Given the limited possibilities for attribution and the absence of any real chance of provoking an armed (or even any kind of) response, ICOs are low-risk, low-cost capabilities that can contribute to the destabilisation of an adversary. If the main goal of an Influence Operations campaign is to sow doubt and confusion in order to undermine trust and confidence in the governments of targeted nations, ICOs can certainly contribute to that. The problematic attribution of cyberattacks ensures that it will generally remain unclear who is actually behind the attack, whilst still allowing for a certain degree of plausible deniability when the source of an attack has been determined. One of the major Influence Operations campaigns we have witnessed during the past few years was the involvement of Russia in the Ukraine crisis. However, it is difficult to determine with certainty if these operations did effectively reach their target audience and were able to achieve their intended effects. What is clear, however, is that both IIOs and ICOs have been used as part of a more or less integrated Influence Operation campaign.

B. The Western approach

In democratic societies, there is almost a firewall between the soft power of IIOs and the hard power of covert or clandestine ICOs. This is not only visible in peacetime, but also during military conflicts. There is, of course, a good reason for this as military Information Operations ‘often involve deception and disinformation that is effective in war but counterproductive in peace’.[29] As described above, this is a distinction that more authoritarian states do not seem to care about as much.

A major drawback of this compartmentalised approach is that it is proving to be very difficult to develop an integrated, national, approach to Influence Operations. In most Western nations and NATO members, strategic thinking about Influence Operations and Information Warfare was principally done by the military, especially after the disintegration of the USSR. Strategic communications have therefore been mostly led by the defence establishment, but in recent years the need for a more comprehensive approach to Influence Operations has begun to be acknowledged at the highest levels of government.[30]

The distinction between soft power and hard power instruments is key to understanding the limitations of the Western approach. As a matter of fact, the core of IIOs for democracies is to

27 Timothy Thomas, Supra, p. 142.
28 David J. Smith. ‘How Russia Harnesses Cyberwarfare’, *Defense Dossier*, Issue 4, 2012, pp. 7-8.
29 Joseph Nye, Public Diplomacy and Soft Power, *The Annals of the American Academy of Political and Social Science* 2008; 616; 94, p. 106.
30 See Paul Cornish, *Strategic Communications and National Strategy*, Chatham House, 2011, p. 4.

tell the truth and act in a manner consistent with its principles. A long term approach built around a carefully developed narrative can only be effective if the facts and messages supporting this narrative are reliable and consistent. Another weakness of this approach is that it assumes a critical thinking and reading ability and interest among the audiences so that the 'most truthful narrative can win' whereas openness to a discourse is a question of faith. The soft power efforts are, by their nature, not only directed at adversary audiences, but also at national audiences and media. This does not mean that Western governments are always truthful in practice, but in theory and as laid down in their policies and strategy documents, this is generally a clearly stated objective.[31]

This compartmentalised approach leaves little room for more clandestine and covert actions, as these will undermine the overall narrative directed at adversary, own and neutral audiences to avoid the risk of undermining one's credibility and narrative. Furthermore, there is a healthy scepticism among populations in democracies that propaganda is not only targeting adversaries but also themselves.[32] Intrusive cyber operations might therefore in the long term do more harm than good by damaging trust among a nation's own population. For democracies, executing coercive Influence Operations, generally, just does not seem to be an option.

In addition, engaging in ICOs that might prove to be intrusive for a state means that it carries out activities that are, in most cases, illegal.[33] Western democracies adhere to the notion that the executive branch of government is bound by domestic laws. This is the fundamental principle of limited government in the legal doctrines of rule of law prevalent in both the common and civil law traditions, and is a vital component of the separation of powers in a democratic regime. This means that the executive can only perform an action if allowed to do so by law. Namely, intelligence services can conduct such operations when, for instance, national security issues are at stake.[34] In that sense, the use of ICOs is rather curbed.

As we have seen, Russia has adopted an integrated approach, which includes ICOs as a tool that can be used in peacetime as well as during an armed conflict. For Western nations, there are solid reasons of transparency, objectivity, and legality to exercise restraint in applying these techniques in a peacetime setting. As a consequence, they have limited the use of ICOs to military operations (MILDEC, CNOs) or specific authorities (primarily intelligence agencies) for very specific purposes.

4. A CLOSER LOOK AT INFLUENCE CYBER OPERATIONS

In this section, we analyse a number of cyber operations whose objective was (or seems to have been) to influence the behaviour of target audiences. It also aims to show how broad the

31 'Maintaining transparency and credibility is paramount in the inform line of effort'. See the US Army Field Manual 3-13 on Inform and Influence Activities, January 2013, p. 2-1.

32 A. R. Pratkanis, 'Winning hearts and minds A social influence analysis', in John Arquilla and Douglas A. Borer (Eds.) *Information Strategy and Warfare A guide to theory and practice*, (New York 2007), p. 78.

33 See the Convention on cybercrime signed on 23 November 2001 and ratified by 48 nations and the Directive 2013/40/EU of the European Parliament and of the Council of 12 August 2013 on attacks against information systems and replacing Council Framework Decision 2005/222/JHA.

34 Operation Cupcake conducted by MI6 in 2011 in Duncan Gardham, 'MI6 attacks al-Qaeda in 'Operation Cupcake'', *The Telegraph*, 2 June 2011, http://www.telegraph.co.uk/news/uknews/terrorism-in-the-uk/8553366/MI6-attacks-al-Qaeda-in-Operation-Cupcake.html.

spectrum of ICOs can be as these techniques are all quite common and might be considered low tech as they can be automated, fairly easily outsourced and by the fact that ready-to-use tools are available online. These different activities do, however, qualify as cyberattacks as defined earlier.

Given their effects, ICOs do not reach the level of an armed attack in the legal sense; that is to say that these activities may not prompt an action in self-defence by the injured state pursuant to article 51 of the United Nations Charter. However, given their low intensity, these attacks do not imply that there is a legal void. In this section, some legal comments are provided concerning the different types of techniques we are looking at in order to come up with potential frameworks. Despite these efforts, we will see that these activities are difficult to grasp through the legal lens.

A. Unauthorised access to an information system

Hacking, or gaining access to a computer system, can enable the attacker to modify data for a particular purpose. Hacking critical information infrastructure can seriously undermine trust in national authorities. For example, in May 2014, the group known as Cyber-Berkut compromised the computers of the Central Election Committee in Ukraine.[35] This attack disabled certain functionalities of the software that was supposed to display real time vote-counting. This hack did not hinder the election process, let alone determine its outcome, as voters had to cast an actual physical ballot. It did, however, damage the credibility of the Ukrainian government in overseeing a fair election process. The impact of this type of attack would obviously have been much greater if it had actually influenced the functioning of the voting system. The attack was carried out by a proxy actor and not directly by the Russian government. Although Cyber-Berkut clearly supports Russian policy towards Ukraine, there is yet no definitive proof that these hacktivists have a direct relationship with Russian authorities.[36] This makes denial of involvement by the Russian government not only plausible, but also irrefutable. From the international law standpoint, the use of proxies to conduct such operations makes it almost impossible to relate these activities to a state actor.

Another example is the security breach that affected the US Office of Personnel Management in 2015. Although this was most likely part of an espionage scheme, it was a major embarrassment for the US government and gave the impression that US authorities were not able to protect sensitive information. As Michael Hayden said, this episode is 'a tremendously big deal, and my deepest emotion is embarrassment'.[37]

B. False flag cyberattacks

In April 2015 the French television network TV5 Monde was the victim of a cyberattack from hackers claiming to have ties with Islamic State's (IS) 'Cyber Caliphate'.[38] TV5 Monde said its TV-station, website, and social media accounts were all hit. In addition, the hackers

35 Nikolay Koval, 'Revolution Hacking', in Kenneth Geers (Ed.), *Cyber War in perspective: Russian Aggression against Ukraine*, NATO CCDCOE Publications, Tallinn, 2015.

36 Tim Maurer, 'Cyber Proxies and the Crises in Ukraine', in Kenneth Geers (Ed.), *Cyber War in perspective: Russian Aggression against Ukraine*, NATO CCDCOE Publications, Tallinn, 2015, p. 85.

37 'Michael Hayden Says U.S. Is Easy Prey for Hackers', *Wall Street Journal*, 21 June 2015, http://www.wsj.com/articles/michael-hayden-says-u-s-is-easy-prey-for-hackers-1434924058.

38 Pierre Haski, 'Des cyberdjihadistes attaquent TV5 Monde : « Puissance inouïe »', Rue89, 9 April 2015, http://rue89.nouvelobs.com/2015/04/09/lattaque-tv5-cyber-djihadiste-dune-ampleur-sans-precedent-258584.

posted documents purporting to be ID cards of relatives of French soldiers involved in anti-IS operations. TV5 Monde regained control over most of its sites after about two hours. In the aftermath of the January 2015 terrorist attacks on Charlie Hebdo, it was quite obvious to the general public and to the investigators that the attackers had ties with the IS organisation.

In June 2015 security experts from FireEye involved in the investigation of the hack revealed that Russian hackers used the pseudonym of IS 'Cyber Caliphate' for this attack. According to them, the Russian hacker group known as APT28 (also known as Pawn Storm, Tsar Team, Fancy Bear and Sednit) may have used the name of IS as a diversionary strategy. The experts noticed a number of similarities in the techniques, tactics, and procedures used in the attack against TV5 Monde and by the Russian group.'[39] This can therefore be qualified as a false flag cyberattack where the use of specific techniques (IP spoofing, fake lines of code in a specific language), will result in misattribution.[40]

Why Russia would hack, or sponsor and condone someone else hacking, a French TV station, and pin the blame on an extremist organisation is unclear, since there seems to be no direct correlation with Russian policies. The only discernible rationale behind these attacks, if conducted by Russia, is to sow confusion and undermine trust in French institutions in a period of national anxiety. TV5 Monde can be blamed for not properly protecting its networks and looking like foolish amateurs, and the French government was seemingly unable to respond in an effective way. Although there is no direct connection, it could be argued that any action that undermined the French government may have led it to act in ways favourable to Russian interests.

Here again, plausible deniability provides enough cover not to worry about the legality of such actions or any response of the victim. The fact that only months later it was discovered that there might be a link to the Russian government highlights the very limited risk of repercussions or countermeasures.

C. DDoS attacks

The most common ICOs are distributed denial of service (DDoS) attacks and these provide a clear illustration of the disruptive effects of ICOs in general. The most famous DDoS attacks were the coordinated ones that occurred in April 2007 in Estonia, during the civil unrest resulting from the government's decision to move a Soviet memorial statue.[41]

DDoS attacks are probably still the prevailing option for many actors, as gaining access to

39 According to FireEye '[t]here are a number of data points here in common […] The 'Cyber Caliphate website', where they posted the data on the TV5 Monde hack was hosted on an IP block which is the same IP block as other known APT28 infrastructure, and used the same server and registrar that APT28 used in the past.' See Pierlugi Paganini, 'FireEye claims Russian APT28 hacked France's TV5Monde Channel', *Security Affairs*, 10 June 2015, http://securityaffairs.co/wordpress/37710/hacking/apt28-hacked-tv5monde.html.

40 We define a false flag attack as 'a diversionary or propaganda tactic of deceiving an adversary into thinking that an operation was carried out by another party'. See Mauno Pihelgas (ed.) *Mitigating Risks arising from False-Flag and No-Flag Cyber Attacks*, https://ccdcoe.org/sites/default/files/multimedia/pdf/False-flag%20and%20no-flag%20-%2020052015.pdf.

41 See Andreas Schmidt, 'The Estonian Cyberattacks', in Jason Healey (Ed.) *A Fierce Domain: Conflict in Cyberspace*, 1986-2012, 2013, published by the Atlantic Council.

a botnet is fairly easy and affordable.[42] DDoS attacks are used to overwhelm the target's resources (degradation) or stop its services (disruption). Attacks only affect the availability of internet services and do not infringe on the confidentiality or integrity of networks and data. The objective of these attacks is, therefore, typically to undermine the targets' credibility.

Although technical solutions exist to mitigate their effects, they are still widely used to embarrass governments or other organisations.[43] In 2014 and 2015, NATO websites were the victims of such a campaign and the disruption prompted significant concern, as the main aim of these attacks was to embarrass and disseminate anti-NATO propaganda and to undermine NATO's readiness to defend itself in cyberspace.[44] They also have a 'paintball effect' as they may give the impression of a severe cyberattack.[45] Last but not least, it is very unlikely that a DDoS attack may be considered as a violation of international law, thus creating grounds for a state to lawfully conduct countermeasures against another state.[46]

D. Website defacements

Although most website defacements or hacks of Twitter accounts have only very limited impact, their results can be quite catastrophic. In 2013 the Twitter account of the Associated Press was hacked and a message claiming the White House was under attack was posted. This sent the stock markets down 1 percent in a matter of seconds. With High Frequency Trading, short interruptions as a result of false messages can have profound financial repercussions.[47]

However, in most cases, website defacements are comparable to graffiti and can be classified as vandalism. Technically, they are not very complicated and, again, the effect lies mainly in the embarrassment it causes to the target. The aim is to sow confusion and undermine trust in institutions by spreading disinformation or embarrass the administrators for poor network defence. The effectiveness of the attack therefore lies in the media reaction;[48] the exposure is far more important than the technical stunt itself. These attacks are minor stings, but taken together they have the potential to erode credibility. Their long term effectiveness, however, is questionable, as people become aware of their limited impact and network security is improved.

42 One of the most used techniques and their number is rising every year. https://www.stateoftheinternet.com/security-cybersecurity-attack-trends-and-statistics.html. 'Attackers can rent DDoS attack services for as little as $5, letting them conduct a few minutes-worth of DDoS attacks against any chosen target' in *The continued rise of DDoS attacks*, Symantec Whitepaper, 21 October 2014, http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/the-continued-rise-of-ddos-attacks.pdf.

43 'DDoS on the Move: in Q1 More Countries Suffered Botnet Attacks, Kaspersky Lab Stats Show', 29 May 2015, http://www.kaspersky.com/about/news/virus/2015/DDoS-on-the-Move-in-Q1-More-Countries-Suffered-Botnet-Attacks-Kaspersky-Lab-Stats-Show.

44 Jeffrey Carr, 'Cyber-Berkut and Anonymous Ukraine: Co-opted Hacktivists and Accidental Comedians', *Digital Dao*, 15 March 2014, http://jeffreycarr.blogspot.ro/2014/03/cyber-berkut-and-anonymous-ukraine-co.html.

45 Thomas Rid and Peter Mc Burney state that 'low-potential cyber weapons resemble paintball guns: they may be mistaken for real weapons, are easily and commercially available, used by many to play and getting hit is highly visible', in Thomas Rid and Peter McBurney, 'Cyber Weapons', *RUSI Journal*, 157:1, 6-13, DOI, http://www.tandfonline.com/doi/abs/10.1080/03071847.2012.664354#.Vrm6V7J95aQ.

46 For legal analysis of the 2007 Cyberattacks on Estonia see Kadri Kaska et al., *International Cyber Incidents; Legal Considerations*, NATO CCDCOE Publications, 2010; Michael Schmitt, 'Below the Threshold' Cyber Operations: The Countermeasures Response Option and International Law, *Virginia Journal of International Law*, Vol. 54:3, 2014.

47 Heidi Moore and Dan Roberts, 'AP Twitter hack causes panic on Wall Street and sends Dow plunging', *The Guardian*, 23 April 2013, http://www.theguardian.com/business/2013/apr/23/ap-tweet-hack-wall-street-freefall.

48 Brian Fung and Andrea Peterson, 'The Centcom hack that wasn't', *The Washington Post*, 12 January 2015, https://www.washingtonpost.com/news/the-switch/wp/2015/01/12/the-centcom-hack-that-wasnt/.

E. Doxing

Another technique that has been widely used in recent years is 'doxing' (or 'doxxing'), which is the practice of revealing and publicising information on an organisation (e.g. Sony Corporation[49]) or an individual (e.g. John Brennan[50]) that is private or classified, so as to publically shame or embarrass targets. There are various ways to obtain this information, ranging from open sources to hacking. This type of action is on the rise and if the data of people like the director of the CIA is accessible, that means that everyone's might be.[51]

Doxing may be used for political purposes. For example, in February 2014, Victoria Nuland, then US Assistant Secretary of State for European and Eurasian Affairs, made a rather obscene comment about the European Union in a telephone conversation with the US Ambassador to Ukraine.[52] This type of incident is embarrassing, but more importantly, can create divisions among allies and jeopardise a common policy to address a crisis situation.

Doxing can be an offshoot of an espionage operation, and thus turned into an ICO. Information obtained through a cyberattack as part of an espionage operation can then be disclosed to undermine the adversary. These activities cannot be qualified as a use of force, or be deemed of a coercive nature under international law.[53]

F. Limited response options

After this short overview, one can see the difficulty in grasping the full implications of these ICOs that span a wide spectrum of activities; from the technically savvy to those that are more content-oriented. The common traits are that they have generally limited impact on the attacked party and their success lies in the response or lack thereof. As a matter of fact, it is difficult to counter an ICO as the course of action to respond to them might actually result in a counterproductive outcome or be disproportionate, and thus lead to escalation.

The international law of state responsibility provides grounds to determine if a state has breached an obligation under customary international law (e.g., violation of sovereignty, violation of the principle of non-intervention) in a way that would be deemed an internationally wrongful act.[54] To identify such a violation, it is necessary to determine whether the actor behind a cyber operation can be linked to a state. In order to achieve that, it is necessary to determine whether that state exercises 'effective control' over the group or organisation in question. According to

49 Kim Zetter, 'Sony got hacked, hard, what we know and don't know so far', *Wired Magazine* 3 December 2014. http://www.wired.com/2014/12/sony-hack-what-we-know/.

50 Sam Thielman, 'High school students hack into CIA director's AOL account', *The Guardian*, 19 October 2015. http://www.theguardian.com/technology/2015/oct/19/cia-director-john-brennan-email-hack-high-school-students.

51 Bruce Schneier, 'The rise of political doxing', *Motherboard*, 28 October 2015, http://motherboard.vice.com/read/the-rise-of-political-doxing.

52 Anne Gearan, 'In recording of U.S. diplomat, blunt talk on Ukraine', *The Washington Post*, 6 February 2014, https://www.washingtonpost.com/world/national-security/in-purported-recording-of-us-diplomat-blunt-talk-on-ukraine/2014/02/06/518240a4-8f4b-11e3-84e1-27626c5ef5fb_story.html.

53 See Russell Buchan, 'The International Legal Regulation of State-Sponsored Cyber Espionage', in Anna-Maria Osula and Henry Rõigas (Eds), *International Cyber Norms, Legal, Policy & Industry Perspectives*, NATO CCDCOE Publications, 2016.

54 See James Crawford, *International Law Commission's Articles on State Responsibility, Introduction, Text and Commentaries*, Cambridge University Press, 2002, The forthcoming Tallinn Manual 2.0 will specifically address the issues of State Responsibility pertaining to cyberspace.

the stringent criteria defined by the International Court of Justice, it is difficult to relate many actions in cyberspace to a state, making the options to respond highly limited.[55]

5. CONCLUSIONS AND RECOMMENDATIONS

The attraction of ICOs for states lies mainly in the fact that they are difficult to attribute, and thus provide a high degree of plausible deniability and limited risk of provoking a strong or quick response from the target nation. However, as we have seen, their scope and applicability are restricted as their impact will generally be limited to harassing and annoying a target audience. In most cases, they are not suited to achieving a profound shift in attitude of a target audience or policy of a nation. Although Russia has embarked on a long term and coordinated IIO campaign against NATO and western democracies, its impact on public opinion is limited and its effectiveness will likely decrease as populations become more aware of Russian intentions and the actual impact of the campaign.[56] This is especially relevant in regard to the ICOs orchestrated by Russia. The more target audiences and organisations become aware of the need for adequate protection of their digital infrastructure and the limited long term impact of cyberattacks, the less useful they will become. Most attacks that can be labelled as ICOs are not highly complex and make use of 'low hanging fruit'; the exploitation of those networks with the weakest defences.

ICOs will, however, remain a nuisance and be able to create a certain amount of confusion. As part of a broader IO campaign, they can fuel an already existing sense of insecurity, and thereby support the overall narrative of the campaign. A study conducted by the Chapman University showed, for instance, that Americans fear a cyber terrorist attack more than a physical terrorist attack.[57] This shows that an adversary can exploit the fear of the unknown, whether that fear is realistic or mostly imaginary.

For NATO members and other democracies, the use of ICOs outside of an armed conflict situation will be limited. As these operations involve intrusive measures, the legal grounds for launching these kinds of attacks are generally lacking. In principle, only the intelligence agencies possess the legal mandate to enter networks in foreign countries, and then only under very specific and supervised conditions.[58] In addition, the importance of transparency of government actions in democracies limits the options for employing covert operations to influence the opinions and attitudes of target audiences as such operations are often associated with PSYOPS or propaganda and thus are frowned upon by public opinion and the media.

55 Michael Schmitt and Liis Vihul, 'Proxy Wars in Cyberspace: The Evolving International Law of Attribution', *Fletcher Security Review*, Vol I, Issue II Spring 2014.

56 'It has been argued that information campaigns and cyber tools at the disposal of Russia have had a significant influence on the crisis in Ukraine. So far no one has convincingly shown the real tangible effects of Russian Information warfare, its army of internet trolls and the use of other cyber-attacks'. See Jyri Raitasalo, 'Hybrid Warfare: where's the beef?' *War on the Rocks*, 23 April 2015, http://warontherocks.com/2015/04/hybrid-warfare-wheres-the-beef/.

57 See the Chapman University Survey on American Fears in 2015, http://www.chapman.edu/wilkinson/research-centers/babbie-center/survey-american-fears.aspx.

58 See for example the FISA (Foreign Intelligence Surveillance Act) in the United States, which provides for strict limitations on foreign surveillance.

To raise awareness, it is also necessary to increase transparency.[59] The media need to be provided with reliable and verifiable information so that the general audience is better informed, and to minimise exaggeration regarding the effects of certain cyberattacks. Additionally, and deriving from this transparency issue, states and corporations will have to learn to deal better in a more transparent and less convulsive way with leaks that are bound to happen, as a secretive and evasive response will merely increase their impact.[60]

In response to ICOs, it is therefore essential that government officials and the public at large have a fundamental grasp of the nature and impact of the multiple kinds of cyberattacks that are possible. They must be aware that hacking the webserver of a TV station does not constitute a serious threat to the security or governability of a nation. Hence, apart from the obvious importance of proper defence of networks and systems, the primary instrument for nations to counter ICOs is to raise cyber awareness among the population at large as well as the bureaucratic and political elite. An important step towards this is to tone down the hyperbole in the media, which is too easily tempted to label everything as 'cyber war'.

59 'Fortunately, the antidote to Netwar poison is active transparency, a function that democracies excel in'. In Robert Brose, 'Cyber War is not Net War, Net War is not Cyber War' in *7th International Conference on Cyber Conflict Proceedings*, NATO CCD COE Publications, 2015, p. 48.

60 Henry Farrell and Martha Finnemore, 'The end of hypocrisy: American Foreign Policy in the Age of Leaks', *Foreign Affairs*, 15 October 2013.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Is the International Law of Cyber Security in Crisis?

Kubo Mačák
Law School
University of Exeter
Exeter, United Kingdom
k.macak@exeter.ac.uk

Abstract: Several indicators suggest that the international law of cyber security is in the midst of a crisis. First, proposals of internationally binding treaties by the leading stakeholders, including Russia and China, have been met with little enthusiasm by other states, and are generally seen as having limited prospects of success. Second, states are extremely reluctant to commit themselves to specific interpretations of the controversial legal questions and thus to express their *opinio juris*. Third, instead of interpreting or developing rules, state representatives seek refuge in the vacuous term 'norms'. This paper argues that the reluctance of states to engage themselves in international law-making has generated a power vacuum, lending credence to claims that international law fails in addressing modern challenges posed by the rapid development of information and communication technologies. In response, a number of non-state-driven norm-making initiatives have sought to fill this vacuum, such as Microsoft's cyber norms proposal or the *Tallinn Manual* project. The paper then contends that this emerging body of non-binding norms presents states with a critical window of opportunity to reclaim a central law-making position, similarly to historical precedents including the development of legal regimes for Antarctica and nuclear safety. Whether the supposed crisis of international law will lead to the demise of inter-state governance of cyberspace or the recalibration of legal approaches will thus be decided in the near future. States should assume a central role in the process if they want to ensure that the existing power vacuum is not exploited in a way that would upset their ability to achieve their strategic and political goals.

Keywords: *attribution, cyber security, governance, international law, international norms, power*

1. INTRODUCTION

None of the global challenges facing the modern international community can be adequately addressed by any single international actor, irrespective of how powerful that actor may be. Whether one thinks of climate change, international terrorism, or cyber threats, all such challenging contemporary phenomena necessitate a framework for international co-operation. It is international law that 'affords [such] a framework, a pattern, a fabric for international society'.[1]

By establishing a framework of constraints, the law simultaneously guarantees a sphere of autonomy for its subjects.[2] In the context of international law, legal norms lay down shared boundaries of acceptable conduct in international relations, while preserving important space for manoeuvre, discretion and negotiation. This is the idea at the root of the famous '*Lotus* presumption',[3] according to which states may generally act freely unless prevented by a contrary rule of international law.[4]

In order to delineate this zone of freedom for states and other international actors with respect to a new phenomenon of international significance, it is necessary to identify, interpret and apply relevant legal rules to it.[5] Cyberspace, broadly understood, is precisely such a phenomenon. Crucially, the uses and abuses of this complex borderless virtual space impinge on vital state interests in the physical world, including national security, public safety, or economic development. As such, cyberspace extends far beyond the domain of internal affairs of any state.[6]

Yet, with respect to the management of cyberspace, it may appear that international law fails to deliver. Although the main building blocks of the Internet's architecture were laid over two decades ago,[7] it took until 2013 for state representatives to agree on the rudimentary threshold assumption that international law actually applies to cyberspace.[8]

Although that agreement was touted at the time as a 'landmark consensus',[9] its actual import is controversial. It was expressed in the form of a non-binding report of a Group of Government

1 Louis Henkin, *How Nations Behave* (2nd edn, Columbia University Press 1978) 5.
2 Cf Joseph Raz, *The Morality of Freedom* (Clarendon Press 1986) 155 ('Autonomy is possible only within a framework of constraints.').
3 See, e.g, James Crawford, *The Creation of States in International Law* (2nd edn, OUP 2006) 41–42 (describing the presumption as a 'part of the hidden grammar of international legal language').
4 PCIJ, *Lotus Case (France v Turkey)* (Merits) [1927] PCIJ Rep Series A No 10, 18.
5 Cf Gennady M Danilenko, *Law-Making in the International Community* (Martinus Nijhoff 1993) 1 (arguing that in order for the international legal system to remain effective, it needs to engage in (1) law-making in novel, so far ungoverned areas and (2) constant upgrading and refinement of the existing law).
6 See also Henry H Perritt, 'The Internet as a Threat to Sovereignty? Thoughts on the Internet's Role in Strengthening National and Global Governance' (1998) 5(2) Indiana Journal of Global Legal Studies 423, 429; Katharina Ziolkowski, *Confidence Building Measures for Cyberspace: Legal Implications* (NATO CCD COE 2013) 165.
7 Tim Berners-Lee, 'Information Management: A Proposal', Internal Memo (CERN, March 1989), <http://cds.cern.ch/record/1405411/files/ARCH-WWW-4-010.pdf>. All Internet resources were accessed on 7 March 2016.
8 Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, UN Doc A/68/98 (24 June 2013) ('GGE Report 2013') 8 [19].
9 United States, Department of State, 'Statement on Consensus Achieved by the UN Group of Governmental Experts On Cyber Issues' (7 June 2013) <http://www.state.gov/r/pa/prs/ps/2013/06/210418.htm>.

Experts (GGE) established by the United Nations (UN) General Assembly.[10] At the time, the group was composed of representatives of 15 UN member states,[11] including the three 'cyber superpowers' China, Russia, and the United States.[12] Its position can thus perhaps be taken as confirming a shared understanding in the international community.[13]

Still, the report poses more questions than it answers. International law is supposed to apply, but *which* international law? Although the group endorsed the centrality of the UN Charter,[14] several of its members have questioned the applicability of a prominent subdomain of international law – the law of armed conflict – to cyber operations.[15] Perhaps more importantly, *how* is international law supposed to apply? It is one thing to know that the online realm is not a lawless world, but quite another to understand how its rules precisely apply to cyber phenomena.[16]

Against this background, this paper examines if the current situation is fairly described as one of crisis. To that end, it weighs three key crisis indicators reverberating around states' general reluctance to engage in law-making in the area of the international law of cyber security[17] (section 2). Since new binding rules are few and far between, it then looks to the pre-existing landscape of international law and the extent to which it provides a regulatory mechanism in its own right (section 3). Subsequently, the paper shows that states' retreat from their traditional legislative role has generated a power vacuum (section 4), triggering a number of non-state initiatives seeking to fill it (section 5). On the basis of historical precedents that include the development of legal regimes for Antarctica and nuclear safety, the paper then argues that states now have a critical window of opportunity to build on the plurality of emerging non-binding norms and thus reclaim their central law-making position (section 6). Whether they succeed in doing so and in what way will determine the answer to the overarching question of this paper.

2. CRISIS INDICATORS

Three indicators of the apparent crisis of international law stand out. First, the area of cyber security appears resistant to codification of the applicable rules in a comprehensive multilateral

10 Ibid.
11 Ibid 12-13.
12 See, e.g, Adam Segal, *The Hacked World Order* (Public Affairs 2016) 40.
13 The UN General Assembly subsequently '[w]elcom[ed]' the GGE report in a unanimously adopted resolution without, however, discussing the details of its contents. See UN GA Res 68/243 (9 January 2014) preambular para 11.
14 GGE Report 2013 (n 8) 8 [19] ('International law, and *in particular the Charter of the United Nations*, is applicable') (emphasis added).
15 See, e.g., US, Office of the Secretary of Defense, *Military and Security Developments Involving the People's Republic of China* (2011) 6 ('China has not yet agreed with the U.S. position that existing mechanisms, such as International Humanitarian Law and the Law of Armed Conflict, apply in cyberspace.'); Elena Chernenko, 'Russia Warns Against NATO Document Legitimizing Cyberwars' *Kommersant-Vlast* (29 May 2013) <http://rbth.com/international/2013/05/29/russia_warns_against_nato_document_legitimizing_cyberwars_26483.html> (reporting the Russian government's scepticism of the *Tallinn Manual*'s endorsement of the applicability of international humanitarian law to cyberspace).
16 See also Anna-Maria Osula and Henry Rõigas, 'Introduction' in Anna-Maria Osula and Henry Rõigas (eds), *International Cyber Norms: Legal, Policy & Industry Perspectives* (NATO CCD COE 2016) 14.
17 The term 'international law of cyber security', as understood in this paper, refers to an emerging legal discipline and a body of law that concerns the rights and obligations of states regarding cyber security.

binding treaty.[18] This is not for want of trying by the leading international stakeholders. Already in 1996, France put forward the earliest proposal with the lofty title *Charter for International Cooperation on the Internet*.[19] Later, a joint Russo-Chinese initiative resulted in two proposals for a *Code of Conduct for Information Security*, submitted to the UN General Assembly in 2011 and 2015, respectively.[20] However, none of these proposals was met with much enthusiasm by other states[21] and scholars describe the prospects of an 'omnibus' treaty being adopted in the near future as slim to negligible.[22]

Second, states have shown extreme reluctance to contribute towards the development of cyber-specific customary international rules. In addition to state practice in this area being inevitably shrouded in secrecy,[23] states have been reluctant to offer clear expressions of *opinio juris* on matters related to cyber security.[24] At times, this approach may certainly be understandable, being the consequence of a domestic political gridlock or even a deliberate waiting strategy.[25] On the whole, however, it adds to the pervasive ambiguity as far as the specific applicability of international law is concerned. This trend is visible even in the most recent developments. A representative example of another missed opportunity to steer the development of cyber custom is provided by the new United States (US) *Law of War Manual* adopted in July 2015.[26] Although it does contain a chapter on cyber operations,[27] the Manual skirts virtually all of the

18 For existing sectoral and regional treaties concerning aspects of cyber security, see text to notes 40–49 below.

19 Timothy S Wu, 'Cyberspace Sovereignty? The Internet and the International System' (1997) 10(3) Harvard Journal of Law & Technology 647, 660. The initiative was reportedly supposed to 'lead to an accord comparable to the international law of the sea, which governs the world's oceans'. 'France Seeks Global Internet Rules', *Reuters News Service* (31 January 1996) <http://dasalte.ccc.de/crd/CRD19960205.html.de>.

20 Letter dated 12 September 2011 from the Permanent Representatives of China, the Russian Federation, Tajikistan and Uzbekistan to the United Nations addressed to the Secretary-General, UN Doc A/66/359 (14 September 2011) 3–5; Letter dated 9 January 2015 from the Permanent Representatives of China, Kazakhstan, Kyrgyzstan, the Russian Federation, Tajikistan and Uzbekistan to the United Nations addressed to the Secretary-General, UN Doc 69/723 (13 January 2015) 3–6.

21 See, e.g., United Kingdom, Response to General Assembly resolution 68/243 'Developments in the field of information and telecommunications in the context of international security' (May 2014) <https://s3.amazonaws.com/unoda-web/wp-content/uploads/2014/07/UK.pdf> 5 (noting that 'attempts to conclude comprehensive multilateral treaties, codes of conduct or similar instruments would [not] make a positive contribution to enhanced international cybersecurity'); Marina Kaljurand, 'United Nations Group of Governmental Experts: The Estonian Perspective' in Osula and Rõigas (n 16) 123 (stating that 'starting negotiations on the draft Code of Conduct … would be premature').

22 See, e.g., Jack Goldsmith, 'Cybersecurity Treaties: A Skeptical View' in Peter Berkowitz (ed), *Future Challenges in National Security and Law* (Hoover Institution Press 2011) 12; Matthew C Waxman, 'Cyber-Attacks and the Use of Force: Back to the Future of Article 2(4)' (2011) 36 Yale Journal of International Law 421, 425–426; Oona A Hathaway et al, 'The Law of Cyber-Attack' (2012) 100 California Law Review 817, 882; Kristen E Eichensehr, 'The Cyber-Law of Nations' (2015) 103 Georgetown Law Journal 317, 356; Michael N Schmitt and Liis Vihul, 'The Nature of International Law Cyber Norms' in Osula and Rõigas (n 16) 39.

23 See Richard A Clarke & Robert Knake, Cyber War: The Next Threat to National Security and What to Do About It (Harper Collins 2010) xi ('The entire phenomenon of cyber war is shrouded in such government secrecy that it makes the Cold War look like a time of openness and transparency.').

24 Notable exceptions include, e.g., US, The White House, *International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World* (2011); Harold Hongju Koh, 'International Law in Cyberspace: Remarks as Prepared for Delivery to the USCYBERCOM Inter-Agency Legal Conference' (18 September 2012) <http://www.state.gov/s/l/releases/remarks/197924.htm>.

25 Michael N Schmitt and Sean Watts, 'The Decline of International Humanitarian Law *Opinio Juris* and the Law of Cyber Warfare' (2015) 50 Texas International Law Journal 189, 211.

26 US, Department of Defense, Office of the General Counsel, *Law of War Manual* (2015) <http://www.dod.mil/dodgc/images/law_war_manual15.pdf>.

27 Ibid ch xvi.

unsettled issues, including standards of attribution, rules of targeting or the requirement to review cyber weapons.[28]

While the first two indicators relate to states' reluctance to act in ways meaningful for the generation of new rules, the third concerns their actual conduct in relation to cyber governance. It would be inaccurate to claim that states have entirely given up on standard-setting. However, instead of interpreting or developing rules of international law, state representatives have sought refuge in the vacuous term 'norms'. We can see this trend most clearly in the context of the work of the UN GGE. In its latest report, the group touted the advantages of '[v]oluntary, *non-binding norms* of responsible state behaviour'.[29] The report claimed that such norms prevent conflict in cyberspace, foster international development, and reduce risks to international peace and security.[30] The report further recommended 11 such norms for consideration by states,[31] while making it clear that these norms operate on a decidedly non-legal plane.[32] Despite their minimalistic nature, the norms have thus far received very limited endorsement by their addressees. For example, at a US-China summit in September 2015, the two participating heads of state 'welcomed' the report but refrained from committing themselves to any of the proposed norms.[33]

Together, these three indicators signify a trend of moving away from the creation of legal rules of international law in the classical sense. Instead of developing binding treaty or customary rules, states resort to normative activity outside the scope of traditional international law. Although this trend appears to be especially prominent in the area of cyber security, it is by no means limited to it. In legal theory, this phenomenon has been described as 'the pluralization of international norm-making',[34] characterised by the observation that 'only a limited part of the exercise of public authority at the international level nowadays materializes itself in the creation of norms which can be considered international legal rules according to a classical understanding of international law'.[35] In order to understand the impact this situation has on the international legal regulation of cyber security, we must zoom out slightly to take in the broader context of existing international law.

3. EXISTING LEGAL LANDSCAPE

The absence of a cyber-specific system of rules of international law does not mean that there are no legal rules that would apply to cyber activities. As we have seen, states accept that generally applicable rules of international law apply to states' conduct in cyberspace, too. This is undoubtedly correct. If international law is to be an efficient governance structure, it must be

28 See further Sean Watts, 'Cyber Law Development and the United States Law of War Manual' in Osula and Rõigas (n 16).

29 Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, UN Doc A/70/174 (22 July 2015) ('GGE Report 2015') 7 [10] (emphasis added).

30 Ibid 7 [10].

31 Ibid 7–8 [13].

32 Ibid 7 [10].

33 US, White House, 'Fact Sheet: President Xi Jinping's Visit to the United States' (25 September 2015) <https://www.whitehouse.gov/the-press-office/2015/09/25/fact-sheet-president-xi-jinpings-state-visit-united-states>.

34 Jean d'Aspremont, *Formalism and the Sources of International Law* (OUP 2011) 222.

35 Ibid 2.

adaptable to new phenomena without the need to reinvent an entire regulation framework on each occasion.[36]

By way of an example, the UN Charter was finalised when the invention of nuclear weapons was still a closely guarded secret and this instrument thus understandably did not refer to this type of weapons in its provisions on the use of force.[37] Still, the International Court of Justice (ICJ) had little difficulty in holding, in the *Nuclear Weapons* Advisory Opinion issued decades later, that those provisions 'apply to any use of force, regardless of the weapons employed',[38] notwithstanding the fact that a particular type of weapons might not yet have been generally known or even invented when the Charter was adopted. Following the same logic, cyber operations must equally be subject to the international law regulation of the use of force.[39]

In addition to these generally applicable rules of international law, certain sectoral and regional treaties taken together provide a 'patchwork of regulations' for cyber activities.[40] These include, in particular, the 1992 Constitution of the International Telecommunication Union;[41] the 2001 Budapest Convention on Cybercrime[42] and its 2006 Protocol on Xenophobia and Racism;[43] the 2009 Shanghai Cooperation Organisation's Information Security Agreement;[44] and the 2014 African Union's Cyber Security Convention.[45] Although important in their own right, these international agreements govern only a small slice of cyber-related activities (such as criminal offences committed by means of computer systems[46] or operations interfering with existing telecommunications networks[47]), or have a very limited membership (six states in the case of the Shanghai Cooperation Organisation's agreement[48] and none yet in that of the African Union's convention[49]).

Therefore, although cyberspace is certainly not a lawless territory beyond the reach of international law, for now there is no complex regulatory mechanism governing state cyber activities.[50] Moreover, states seem reluctant to engage themselves in the development and

36 Cf US, *International Strategy for Cyberspace* (n 24) 9.
37 Charter of the United Nations (signed 26 June 1945, entered into force 24 October 1945) 1 UNTS 16, Arts 2(4) and 39–51.
38 ICJ, *Legality of the Threat or Use of Nuclear Weapons Case* (Advisory Opinion) [1996] ICJ Rep 226 [39].
39 Accord Michael N Schmitt (ed), *Tallinn Manual on the International Law Applicable to Cyber Warfare* (CUP 2013) 42.
40 Hathaway (n 22) 873.
41 Constitution of the International Telecommunication Union (concluded 22 December 1992, entered into force 1 July 1994) 1825 UNTS 143 (hereinafter 'ITU Constitution').
42 Council of Europe, Convention on Cybercrime (signed 23 November 2001, entered into force 1 July 2004) ETS 185.
43 Council of Europe, Additional Protocol concerning the Criminalisation of Acts of a Racist and Xenophobic Nature Committed through Computer Systems (opened for signature 28 January 2003, entered into force 1 March 2006) ETS 189.
44 Agreement between the Governments of the Member States of the Shanghai Cooperation Organisation on Cooperation in the Field of International Information Security (signed 16 June 2009, entered into force 5 January 2012) ('Yekaterinburg Agreement').
45 African Union Convention on Cyber Security and Personal Data Protection (signed 27 June 2014) AU Doc EX.CL/846(XXV).
46 Convention on Cybercrime (n 42) Arts 2–10.
47 ITU Constitution (n 41) Art 45 (prohibiting harmful interference) and Annex (defining harmful interference).
48 Yekaterinburg Agreement (n 44).
49 See further Henry Rõigas, 'Mixed Feedback on the "African Union Convention on Cyber Security and Personal Data Protection"', *CCD COE INCYDER Database* (20 February 2015) <https://ccdcoe.org/mixed-feedback-african-union-convention-cyber-security-and-personal-data-protection.html>.
50 See also Hathaway (n 22) 873.

interpretation of international law applicable to cyber security. This voluntary retreat has generated a power vacuum, enabling non-state actors to move into the space vacated by states and pursue various forms of 'norm entrepreneurship'.[51]

4. POWER VACUUM

Vectors of power and law do not overlap perfectly. State power is certainly influenced by many other factors, which may include military might, wealth, and moral authority.[52] Nonetheless, it needs little emphasis that the powerful normally seek to use legal regulation to consolidate and project their power.[53] If we understand power simply as 'the ability to alter others' behaviour to produce preferred outcomes',[54] then setting legal obligations is one way how to exercise this ability. Everything else being equal, it is more likely than not that these 'others' will act in accordance with a certain standard of behaviour when it is required by law than when it is not.

Yet, legal uncertainty may at times be deemed desirable by even the most powerful states. For example, during the early days of space exploration, only two states were capable of acting in outer space: the US and the Soviet Union. Yet these two states resisted, for a significant time, to commit themselves to any binding rules that would govern outer space. Both believed that the adoption of such rules would only serve to constrain their activities in space. In that vein, '[l]egal uncertainty was useful to those with the power to act in space, on either side of the cold war.'[55]

However, cyberspace and outer space – albeit frequently lumped together as so-called 'global commons'[56] – are decidedly different from one another. This is not only because many states are challenging the very idea of cyberspace as commons by seeking to assert greater control online.[57] More importantly, cyberspace is already a much more crowded domain than outer space could ever be. To wit, the US and the Soviet Union were not just the only *states* engaged in space exploration for several decades, they were also the only *actors* capable of space flight as such. In contrast, cyberspace is populated primarily by non-state actors, which include individuals, corporations, and other more loosely organised groups.[58] The possibility of anonymity online combined with the corresponding difficulty of attribution of cyber operations

51 Martha Finnemore and Kathryn Sikkin, 'International Norm Dynamics and Political Change' (1998) 52(4) International Organization 887.

52 Michael Byers, *Custom, Power and the Power of Rules* (CUP 1999) 5.

53 See further Rosalyn Higgins, *Problems and Process: International Law and How We Use It* (Clarendon Press 1995) 3–4 (analysing the relationship between law and power from the perspective of international law).

54 Joseph Nye, *The Future of Power* (Public Affairs 2011) 10.

55 Stuart Banner, *Who Owns the Sky? The Struggle to Control Airspace from the Wright Brothers On* (Harvard University Press 2008) 278.

56 See, e.g., Mark Barrett et al, *Assured Access to the Global Commons* (NATO 2011) xii; Scott Jasper and Scott Moreland, 'Introduction' in Scott Jasper (ed), *Conflict and Cooperation in the Global Commons* (Georgetown University Press 2012) 21; Nicholas Tsagourias, 'The Legal Status of Cyberspace' in Nicholas Tsagourias & Russell Buchan, *Research Handbook on International Law and Cyberspace* (Edward Elgar 2015) 24–25; Paul Meyer, 'Outer Space and Cyberspace: A Tale of Two Security Realms' in Osula and Rõigas (n 16) 157.

57 Scott Shackelford, *Managing Cyber Attacks in International Law, Business, and Relations* (CUP 2014) 58.

58 See further Johan Sigholm, 'Non-State Actors in Cyberspace Operations' (2013) 4(1) Journal of Military Studies 1, 9–23.

have resulted in the 'dramatic amplification' of power in the hands of these non-state actors at the expense of their state counterparts.[59]

The effect of legal uncertainty is thus much more complex than what we saw in relation to outer space, as it affects a far more populous spectrum of actors, state and non-state alike. Accordingly, non-state actors have now moved into the vacated norm-creating territory previously occupied exclusively by states. These developments have been primarily driven by the private sector and by the academia, as epitomised by Microsoft's cyber norms proposal and by the so-called *Tallinn Manual* project.

5. NON-STATE-DRIVEN INITIATIVES

The more recent of the two, Microsoft's proposal entitled International *Cybersecurity Norms: Reducing Conflict in an Internet-Dependent World* was published in December 2014.[60] Interestingly, this was not the first private-sector initiative of this kind. Exactly 15 years earlier, Steve Case, then the CEO of AOL, urged states to revise their 'country-centric' laws and adopt instead 'international standards' governing crucial aspects of conduct online, including security, privacy, and taxation.[61] Still, Microsoft's text is the first comprehensive proposal of specific standards of behaviour online, which, despite its private origin, proposes norms purporting to regulate solely the conduct of states.[62] The openly proclaimed central aim of this white paper was to reduce the possibility that ICT products and services would be 'used, abused or exploited by nation states as part of military operations'.[63] To that end, the paper put forward six cyber security norms, which collectively called on states to improve their cyber defences and limit their engagement in offensive operations.[64]

In 2013, an international group of experts led by Professor Michael Schmitt published the *Tallinn Manual on the International Law Applicable to Cyber Warfare*.[65] Although the project was undertaken under the auspices of the Estonia-based NATO Cooperative Cyber Defence Centre of Excellence (CCD COE), the *Manual* makes it clear that its text should be seen as reflecting the views of the experts themselves and not the states or institutions from which they originated.[66] As apparent from its title, the *Manual* maintains a clear military paradigm throughout, focussing on the law on the use of force (*jus ad bellum*) and the law of armed conflict (*jus in bello*).[67] Its text identifies 95 rules adopted by consensus among the group of experts who

59 Christian Czosseck, 'State Actors and their Proxies in Cyberspace' in Katharina Ziolkowski (ed), *Peacetime Regime for state Activities in Cyberspace* (NATO CCD COE 2013) 1–3.

60 Angela McKay et al, *International Cybersecurity Norms: Reducing Conflict in an Internet-Dependent World* (Microsoft 2014) <http://aka.ms/cybernorms>.

61 Steve Case, 'Remarks Prepared for Delivery (via satellite) Israel '99 Business Conference' (13 December 1999), cited in Jack Goldsmith and Timothy S Wu, *Who Controls the Internet?: Illusions of a Borderless World* (OUP 2006) 194.

62 McKay et al (n 60) 2–3.

63 Suzanne Choney, '6 Proposed Cybersecurity Norms Could Reduce Conflict', *Microsoft: The Fire Hose* (5 December 2014) <https://blogs.microsoft.com/firehose/2014/12/05/6-proposed-cybersecurity-norms-could-reduce-conflict>.

64 McKay et al (n 60) 2. The complete list of the proposed norms may be found in the annex to the document: ibid 20.

65 *Tallinn Manual* (n 39).

66 Ibid 11.

67 Ibid 5.

were guided by the ambition to 'replicate customary international law'.[68] Early reviews of the *Manual* criticised its almost exclusive focus on activities occurring above the level of the use of force, whereas in reality, most (if not all) cyber operations fall below that threshold.[69] However, the ongoing '*Tallinn 2.0*' project, scheduled for completion in 2016, should dispel some of these objections by turning its attention to 'below-the-threshold' operations and by addressing issues of state responsibility, the law of the sea, international telecommunications law, and even human rights law.[70] Like the Microsoft paper, both iterations of the *Tallinn Manual* project put forward standards of state behaviour and are avowedly state-centric in their approach.

Understandably, the two initiatives differ in important ways. The 'norms' proposed by Microsoft are clearly meant as broad suggestions only, meaning that states need to transform them into more specific commitments. For instance, norm 2 stipulates that 'states should have a clear principle-based policy for handling product and service vulnerabilities that reflects a strong mandate to report them to vendors rather than to stockpile, buy, sell, or exploit them'.[71] As recognised in the paper itself, such policies need to be developed by each individual state and tailored to the needs of the concerned state.[72]

By contrast, the *Tallinn Manual* 'rules' take on the more restrictive and specific form of purported customary legal obligations, which should simply be observed by states as binding without the need for their further endorsement or adaptation.[73] In other words, the *Manual* aims to interpret how 'extant legal norms' apply to conduct in cyberspace,[74] and not to 'set forth *lex ferenda*'.[75] Yet, given that the *Manual* frequently puts forward detailed and novel positions, it does not always succeed in maintaining a bright line between norm interpretation and norm development.[76] Nevertheless, the purported rules it contains are much more specific than Microsoft's cybersecurity norms. For example, rule 37 sets out the prohibition of cyber attacks against civilian objects in the context of an armed conflict.[77] Both crucial terms – 'cyber attacks' as well as 'civilian objects' – are precisely defined by the *Manual*.[78] Although some disagreements may persist about the application of the rule in particular circumstances,[79] the content of the norm is sufficiently clear and precise to generate legal rights and obligations.

However, what initiatives like Microsoft's white paper or the *Tallinn Manual* project share is their non-state origin and expressly non-binding nature. Microsoft was keenly aware of its proposal's

68 Ibid 6.
69 See, e.g., Dieter Fleck, 'Searching for International Rules Applicable to Cyber Warfare: A Critical First Assessment of the New Tallinn Manual' (2013) 18(2) Journal of Conflict and Security Law 331, 332–335; Eichensehr (n 22) 589.
70 See 'Tallinn Manual', NATO CCD COE (undated) <https://ccdcoe.org/research.html>.
71 McKay et al (n 60) 12.
72 Ibid.
73 *Tallinn Manual* (n 39) 6.
74 Ibid 1.
75 Ibid 5.
76 See further Kubo Mačák, 'Military Objectives 2.0: The Case for Interpreting Computer Data as Objects under International Humanitarian Law' (2015) 48(1) Israel Law Review 55, 59–63 (discussing the distinction between *lex lata* and *lex ferenda* in the *Tallinn Manual*).
77 *Tallinn Manual* (n 39) 124.
78 Ibid 91 (definition of cyber attack) and 125 [3] (definition of civilian objects).
79 See, e.g., the debate whether computer data may constitute an 'object' for the purposes of international humanitarian law: Heather A Harrison Dinniss, 'The Nature of Objects: Targeting Networks and the Challenge of Defining Cyber Military Objectives' (2015) 48(1) Israel Law Review 39; Mačák (n 76); Michael N Schmitt, 'The Notion of 'Objects' during Cyber Operations: A Riposte in Defence of Interpretive and Applicative Precision' (2015) 48(1) Israel Law Review 81.

limitations in this respect and noted that it merely 'encouraged' states to set the proposed norms on the trajectory towards making them first 'politically' and then 'legally' binding.[80] Similarly, the *Manual* noted in its opening pages that it was meant to be 'a non-binding document'.[81] As the texts in question are in their entirety the products of non-state initiatives, they could hardly amount to anything else. After all, with potential minor qualifications in the area of collective security, it is still true that only 'the states are the legislators of the international legal system'.[82]

If these texts are non-binding, one might question their relevance from the perspective of international law altogether. True, their normativity (in the sense of the strength of their claim to authority[83]) is lower than that of international legal rules. But that does not mean that these efforts are wholly irrelevant for the formation of rules of international law, and even less do they document any supposed irrelevance of international law to the area of cyber security. On the contrary, non-state-driven initiatives of this kind potentially amount to 'a vital intermediate stage towards a more rigorously binding system, permitting experiment and rapid modification'.[84] Moreover, they render the law-making process more multilateral and inclusive than the traditional state-driven norm-making can ever be.[85] Therefore, the crucial question is whether states decide to pick up the gauntlet thrown at them by their non-state counterparts and reclaim their role as principal lawmakers.

6. STATES AT A CRITICAL JUNCTURE

The current situation is certainly not without prior historical parallels. Cyberspace is not the first novel phenomenon to have resisted the development of global governance structures for some time after its emergence. A degree of waiting or stalling may even reflect states' desire to obtain a better understanding of the new phenomenon's strategic potential.[86] Yet with states' improved comprehension of the new situation, their willingness to subject themselves to binding rules usually increases, too. Even the domain of outer space has been eventually subjected to a binding legal regime,[87] despite the strong initial reluctance of the dominant spacefaring states.[88]

Other domains with a higher number of participants may provide more appropriate analogies. For instance, in the context of Antarctica, many non-binding norms were put forward in the 1960s and 1970s with the aim to conserve living and non-living resources of the Antarctic

80 McKay et al (n 60) 3.
81 *Tallinn Manual* (n 39) 1.
82 Stefan Talmon, 'The Security Council as World Legislature' (2005) 99 AJIL 175, 175. As the title of Professor Talmon's article suggests, the qualification to that general observation arises from the Security Council's recent practice of adopting resolutions containing obligations of general and abstract character.
83 Samantha Besson, 'Theorizing the Sources of International Law' in Samantha Besson & John Tasioulas (eds), *The Philosophy of International Law* (OUP 2010) 173.
84 Hugh Thirlway, *The Sources of International Law* (OUP 2014) 164, paraphrasing Mary E O'Connell, 'The Role of Soft law in a Global Order' in Dinah Shelton (ed), *Commitment and Compliance: The Role of Non-Binding Norms in the International Legal System* (OUP 2000) 100.
85 Besson (n 83) 170–171.
86 Cf Patrick W Franzese, 'Sovereignty in Cyberspace: Can It Exist?' (2009) 64 Air Force Law Review 1, 38; Schmitt and Vihul (n 22) 38.
87 See, principally, Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies, 610 UNTS 205 (opened for signature 27 January 1967, entered into force 10 October 1967) ('Outer Space Treaty').
88 See text to n 55 above.

environment.[89] These norms gradually evolved into the 1991 Antarctic Environmental Protection Protocol, a complex binding instrument that has since been ratified by all key stakeholders.[90]

Similarly, it took over three decades since the 1954 launch of the first nuclear power plant in the world in Obninsk, Soviet Union,[91] until the first international conventions on nuclear safety were adopted.[92] In the meantime, states were guided by non-binding safety standards and criteria, most of which were issued by the International Atomic Energy Agency (IAEA).[93] Afterwards, nuclear safety conventions consolidated this emerging body of non-binding norms and made many of the relevant standards mandatory for all member states.[94]

As these examples demonstrate, instead of lamenting over a supposed crisis of international law, it is more appropriate to view the current situation as an intermediate stage on the way towards the generation of cyber 'hard law'. Non-state-driven initiatives provide opportunities for states to identify overlaps with their strategic interests and they may serve as norm-making laboratories. Their usefulness in this sense is confirmed by a recent report of the EastWest Institute, which helpfully maps out areas of convergence across various proposals of norms of state behaviour in cyberspace including those analysed in this paper.[95]

A final point to consider is the so-called attribution problem (understood as the difficulty in determining the identity or location of a cyber attacker or their intermediary[96]). For some time, it was rightly seen as an impediment to the development of effective legal regulation of cyber activities. It was argued that the prevailing anonymity online 'makes it difficult – if not impossible – for rules on either cybercrime or cyberwar to regulate or deter.'[97] However, recent technological progress has translated into increased confidence of states with respect to attribution of cyber activities. For instance, the US has claimed that it now has the capacity

89 Christopher C Joyner, 'The Legal Status and Effect of Antarctic Recommended Measures' in Dinah Shelton (ed), *Commitment and Compliance: The Role of Non-binding Norms in the International Legal System* (OUP 2003) 175–176.

90 Protocol on Environmental Protection to the Antarctic Treaty (signed 1991, entered into force 14 January 1998) 30 ILM 1455.

91 Paul R Josephson, *Red Atom: Russia's Nuclear Power Program from Stalin to Today* (University of Pittsburgh Press 2005) 2.

92 Convention on Early Notification of a Nuclear Accident (adopted 26 September 1986, entered into force 27 October 1986) 1439 UNTS 275; Convention on Assistance in the Case of a Nuclear Accident or Radiological Emergency (adopted 26 September 1986, entered into force 26 February 1987) 1457 UNTS 133. Two additional conventions were adopted in the 1990s: Convention on Nuclear Safety (done 20 September 1994, entered into force 24 October 1996) 1963 UNTS 293; Joint Convention on the Safety of Spent Fuel Management and on the Safety of Radioactive Waste Management (signed 29 September 1997, entered into force 18 June 2001) (1997) 36 ILM 1436.

93 For an overview of these standards, see IAEA, 'Measures to Strengthen International Co-operation in Nuclear, Radiation, Transport and Waste Safety', IAEA Doc GC(45)/INF/3, Attachment 2 (31 August 2001) 1–7.

94 See further Norbert Pelzer, 'Learning the Hard Way: Did the Lessons Taught by the Chernobyl Nuclear Accident Contribute to Improving Nuclear Law?' in the Joint Report by the OECD Nuclear Energy Agency and the International Atomic Energy Agency, *International Nuclear Law in the Post-Chernobyl Period* (OECD 2006) 86–88.

95 Greg Austin, Bruce McConnell, and Jan Neutze, 'Promoting International Cyber Norms: A New Advocacy Forum' (EastWest Institute, December 2015) <http://issuu.com/ewipublications/docs/bgcybernorms> 10–17.

96 David A Wheeler and Gregory N Larsen, 'Techniques for Cyber Attack Attribution' (Institute for Defense Analysis 2003) 1.

97 Duncan B Hollis, 'An e-SOS for Cyberspace' (2011) 52(2) Harvard International Law Journal 374, 378.

to locate its cyber adversaries and hold them accountable.[98] In a similar statement, Canada noted that it has robust systems in place allowing it to localise cyber intrusions, including those orchestrated by state-sponsored actors.[99] Significant progress has also been made in the understanding of the legal standards of attribution as applied to online conduct.[100] Although it is probably correct that the attribution problem can at most be managed but not solved,[101] these developments show that time may be ripe for states to endorse the regulatory and deterrent potential of international legal rules.

Building on the emerging normative convergence identified above, states should be able to reclaim their central role in international law-making. In the more immediate future, they should become more forthcoming in expressing their opinion as to the interpretation of existing international law to cyber issues.[102] This will in time enable the applicable *opinio juris* to consolidate, thus facilitating the process of transformation of state power into obligations of customary law.[103] Additionally, states should gradually overcome their current aversion to treaty commitments. Reports from late 2015 that the US and China started negotiating a binding arms control treaty for cyberspace are possible early signs that this process is already underway.[104] Finally, this iterative process of state-appropriated norm-making could in the long run quite plausibly result in the adoption of one or several comprehensive multilateral undertakings, possibly commencing with definitional matters to pave the way towards future consensus-building over more substantive issues.[105]

7. CONCLUSION

International law of cyber security is at a critical juncture today. It is true that states' hesitation to engage in the development and application of international law has generated a power vacuum allowing for the emergence of non-state norm-making initiatives. Still, it would be premature to speak of a situation of crisis.

Several historical parallels show that a mixture of initial soft-law approaches combined with a growing set of binding rules can provide a logical and functioning response to a novel

98 Zachary Fryer-Biggs, 'DoD's New Cyber Doctrine: Panetta Defines Deterrence, Preemption Strategy', *Defense News* (13 October 2012) http://archive.defensenews.com/article/20121013/DEFREG02/310130001/DoD-8217-s-New-Cyber-Doctrine" http://archive.defensenews.com/article/20121013/DEFREG02/310130001/DoD-8217-s-New-Cyber-Doctrine.

99 Canada, Statement by the Chief Information Officer for the Government of Canada (29 July 2014) <http://news.gc.ca/web/article-en.do?nid=871449>.

100 See, e.g., Nicholas Tsagourias, 'Cyber Attacks, Self-Defence and the Problem of Attribution' (2012) 17(2) Journal of Conflict and Security Law 229; Zhixiong Huang, 'The Attribution Rules in ILC's Articles on State Responsibility: A Preliminary Assessment on Their Application to Cyber Operations' (2015) 14 Baltic Yearbook of International Law 41.

101 Thomas Rid and Ben Buchanan, 'Attributing Cyber Attacks' (2014) 38 Journal of Strategic Studies 1, 28.

102 For other similar calls on states to be more proactive in expressing their cyber-specific *opinio juris*, see, e.g., Katharina Ziolkowski, 'General Principles of International Law as Applicable in Cyberspace' in Ziolkowski (n 59) 175; Schmitt and Vihul (n 22) 47; Schmitt and Watts (n 25) 230–231.

103 Cf Byers (n 52) 18.

104 David E Sanger, 'U.S. and China Seek Arms Deal for Cyberspace', New York Times (19 September 2015) <http://www.nytimes.com/2015/09/20/world/asia/us-and-china-seek-arms-deal-for-cyberspace.html?_r=0>.

105 See, e.g., Hathaway (n 22) 877.

phenomenon. In the 21st century, pluralisation of norm-making processes involving diverse state and non-state actors is a common feature at the international level and it need not be feared as such.[106]

What matters is whether, and to what extent, states will reclaim their traditional central legislative role. Their conduct in the next few years will determine whether we will observe a gradual demise of inter-State governance of cyberspace or a fundamental recalibration of legal approaches with states taking centre stage once again. If they want to ensure that the existing power vacuum is not exploited in a way that might upset their ability to achieve their strategic and political goals, states should certainly not hesitate too long.

ACKNOWLEDGMENTS

I am grateful to Louise Arimatsu, Ana Beduschi, Russell Buchan, Michael N. Schmitt, and the anonymous reviewers for their valuable comments and suggestions on earlier drafts of this paper. Any remaining errors or omissions are my own responsibility.

106 See d'Aspremont (n 34) 2–3.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Conceptualising Cyber Arms Races

Anthony Craig
Cardiff University
Cardiff, United Kingdom

Dr Brandon Valeriano
Cardiff University
Cardiff, United Kingdom

Abstract: This paper investigates the emergence of an arms race dynamic in the international cyber domain. The numerous claims made of an ongoing cyber arms race by the media and other analysts have not been backed up by careful empirical analysis. Characterised by the competitive and rapid mutual build-up of capabilities between pairs of states, arms races are a long standing aspect of study in international relations, with statistical evidence suggesting a relationship between these factors and war. Our work extends the tradition of arms race scholarship to the field of cyber security by providing a methodology for accounting for the build-up of cyber capabilities by nation states. We examine the concept of the cyber arms race and provide a plausibility probe for a macro study by examining the cases of the United States and Iran, and of North Korea and South Korea. We employ time series data on a number of indicators to measure each state's scale of increase in cyber capabilities, before investigating whether the states in question are directing their efforts against one another. Our findings suggest that these state dyads have indeed been engaged in cyber arms races, as defined by their competitive and above-normal mutual increase in cyber capabilities. This work furthers our understanding of state behaviour in the cyber domain, and our methodology helps to establish a pathway for the future extensive data collection of this new phenomenon.

Keywords: *cyber conflict, arms race, cyber capabilities*

1. INTRODUCTION

Cyberspace is now considered the fifth domain of warfare after land, sea, air, and space. Cyber conflict can be defined as 'the use of computational technologies in cyberspace for malevolent and/or destructive purposes in order to impact, change, or modify diplomatic and military interactions between entities' (Valeriano and Maness, 2015, p.32). It consistently tops national threat assessments by policy figures, and in 2012 the US Defence Secretary warned of a 'cyber Pearl Harbor' that could devastate the country's critical infrastructure (Bumiller and Shanker, 2012). Regardless of the accuracy of these statements, there is a growing understanding that

such insecurities are driving countries to channel the ever increasing resources into their ability to defend themselves against cyber actions, and to launch offensive operations.

Media reports frequently use the term 'arms race' to describe the global proliferation of cyber warfare capabilities as states respond to their security concerns (Corera, 2015). For 57% of security experts and policy elites, the cyber arms race is a reality according to a 2012 survey (McAfee, 2012). Arms races traditionally refer to the rapid threat-driven and competitive build-up of military power between two countries, and have been criticised in the study of international relations due to their escalatory potential in bringing countries closer to the brink of war. Yet until now, the idea of a cyber arms race has not been subjected to proper empirical and academic analysis.

Here we conceptualise cyber arms races by applying traditional arms race theory to the cyber domain, thus gaining important insights into one of the most pressing and rapidly developing issues in world politics. First, the arms race and cyber literature is consulted before setting out our methods, and then two case studies of international rival state pairs are presented: the United States and Iran, and North and South Korea. We first measure their scale of arming, and judge whether it represents abnormal rates of increase. Then we investigate the extent to which these build-ups in cyber power occur in competition with one another specifically. We conclude by discussing the implications of our findings for interstate cyber relations, as well as their limitations, and explain how our research paves the way for future quantitative research on cyber capabilities.

2. WHAT IS AN ARMS RACE?

Arms races have been the subject of much research in the field of international relations as scholars have attempted to investigate their causes and consequences. In its traditional conceptualisation, an arms race results from mutual insecurity and the need to defend against an external threat. The build-up of arms is a core principle in realist theory, which tells us that the anarchical and self-help nature of the international system creates powerful incentives for countries to seek security through military strength and deter potential aggressors in an environment where they can never trust others' intentions.

Rather than promote stability, however, military build-ups can give rise to a security dilemma whereby 'many of the means by which a state tries to increase its security decrease the security of others' (Jervis, 1978). Security-seeking actions are often perceived as threatening and met with reactions in kind, causing interstate tensions to spiral out of control. Decades of peace science research has shown that arms races are associated with an increased likelihood of conflict, whereas very little evidence has been found in support of the opposing deterrence and balance of power theories (Leeds and Morgan, 2012, p.144).

Richardson (1960) made one of the first attempts at mathematically modelling this action-reaction dynamic, and in his set of equations each state's rate of arming increases in response to increases in its rival's military spending. This understanding of the arms race is one of mutual

fear, although Glaser (2004) notes how arms competition can occur when a status quo actor seeks to deter a power-seeking revisionist actor whose motivations are not those of insecurity. Psychology plays an important role in the arms racing process as policy makers do not always act rationally or with complete information (Jervis, 1976). Rather than react to actual threats, the decision to arm is often based on the 'subjective interpretations of the actions of others' (Hammond, 1993, p.47). The response to threats is therefore as much about perceptions as it is about reality.

A distinction can also be made between the types of capabilities involved. Qualitative arms races refer to the competition over technological advances in weaponry, whereas a quantitative arms race is the competition over the sheer number of military forces (Huntington, 1958). When measuring arms races using military expenditures it is important to note that a qualitative improvement in military capability will not necessarily be reflected in a state's military expenditure levels since new and improved weapons systems may be procured less cost (Valeriano, Sample, and Kang, 2013).

Gray (1971, p.40) provides a useful definition of the arms race as:

> 'two or more parties perceiving themselves to be in an adversary relationship, who are increasing or improving their armaments at a rapid rate and structuring their respective military postures with a general attention to the past, current, and anticipated military and political behaviour of the other parties'.

It seems that any form of 'race' in military capabilities should fundamentally feature abnormally high rates of arming by at least two states which are engaged in this behaviour with reference to, and in competition with, one another.

Identifying such a process requires a distinction be made between normal and abnormal rates of military increase. One method used frequently in large-N studies (Sample, 1997; Gibler et al., 2005) codes a rapid build-up if a state's annual growth in either military expenditure or personnel reaches 8% in each of three consecutive years. An alternative measure by Horn (1987) posits that a state is engaged in a rapid military build-up in a given year if its average growth rate in expenditure in the preceding five years is greater than that of the preceding ten years; and if this ten year average is greater than that of the entire time period under observation. Overall, the current lack of data in this relatively new and often secretive domain means that alternative methods for evaluating the magnitude of cyber build-ups will need to be used.

In these quantitative studies, the competitive aspect is also measured in various ways. Sample (1997) uses data on militarised interstate disputes (the threat, display, or use of force) to confirm an adversarial relationship. Gibler et al. (2005) code their arms races based on Thompson's (2001) dataset of ongoing rivalries. In qualitative studies such as this, however, a more in depth analysis of the dyadic relationship can help uncover an action-reaction dynamic.

3. THE CYBER DOMAIN

Cyberspace is defined by Nye (2011, p.19) as the 'Internet of networked computers but also intranets, cellular technologies, fibre optic cables, and space based communications'. Cyberspace refers to not only 'all of the computer networks in the world' but also to 'everything they connect and control' (Clarke and Knake 2010, p.70), highlighting the potential risk to a nation's infrastructure given the fact that these systems are often dependent on Internet networks.

According to Choucri (2010, p.228), the development of cyberspace has put states in an 'unprecedented situation' characterised by high levels of uncertainty as they try to maintain control in the face of a changing global security environment. Proponents of the 'cyber-revolution' hypothesis highlight the serious damage cyber conflict could inflict potentially, and in doing so elevate the threat to the top of the state's national security concerns (Clarke and Knake, 2010; Kello, 2013). Others argue to the contrary that the threat is inflated and disconnected from reality (Lindsay, 2013; Valeriano and Maness, 2015), and as we know from traditional arms racing, fear and perceptions can be just as powerful drivers of security competition as actual threats.

Several characteristics help to establish the perception of cyberspace as an inherently insecure environment. Cyber weapons are essentially 'computer codes' used to inflict harm (Rid and McBurney, 2012, p.6), meaning that unlike the physical warfare domain, the virtual nature of malware makes it very difficult for states to gain an accurate picture of one another's capabilities. The anonymity that cyber methods can provide the attacker and the resulting attribution problem add to this uncertainty. Cyber capabilities include the malicious code created as well as the units mobilised, and the hardware and software developed, to defend against such code. Since cyber technologies can be much cheaper than conventional weapons, weaker states can possibly gain asymmetric advantages by entering into the cyber arms arena and compete on a more even footing with traditionally powerful states. The sources of threat are therefore potentially more widespread.

The belief that the cyber conflict domain favours the offense also creates insecurity. The offense-defence balance theory postulates that if offensive military capabilities hold advantages over defensive capabilities, the security dilemma is more intense and the risk of arms races and war greater (Glaser and Kaufmann, 1998, p.47). Offensive cyber capabilities are assumed to be more cost effective and efficient, whereas defence is difficult given the immense challenge involved in securing every civilian and privately owned network and closing every vulnerability, many of which go undetected until an attack has pointed them out (Liff, 2012). The Internet's lack of geographical constraints further undermines the utility of defence. Offensive preparations may therefore become the dominant strategy, which can risk setting off the security dilemma.

Given the complexity of the cyber domain and its overall novelty, many make statements about the dynamics of cyber conflict without clear connections to more than a few cases, which may be outliers. This is why a macro and empirical perspective on cyber arms build-ups is an important task in the field. Exploring the concept of the cyber arms race is a theoretically appropriate

undertaking given the heightened perceptions of threat that characterise the international cyber domain, and will help shed light on how states are reacting to their cyber security concerns.

4. METHODOLOGY

Since cyber arms races are as yet an untested phenomenon, this study can be regarded as a 'plausibility probe' (Eckstein, 1975) to help decide whether the concept shows promise in application. We conduct case study analyses of two rival state dyads; the United States and Iran, and North and South Korea. These four countries are among the major players in the cyber conflict arena, and are therefore of great interest to policy makers and academics alike.

The 'structured and focused' case study design (George and Bennett, 2005) is adopted here to identify the presence of cyber arms racing behaviour. This approach structures the analysis by asking similar questions of each case, and focuses on the key aspects of the dyadic relationship that will engage the research question. The two questions asked ensure that the essential arms race criteria are met:

1. Are both states engaged in a rapid build-up of cyber capabilities?
2. Are the states in competition with one another?

To answer the first question, time series data is presented to track changes in each state's cyber capabilities. Clearly it is not possible to quantify the actual cyber 'arms' or malware possessed by states, and we acknowledge this limitation. Instead, our approach is inspired by the Correlates of War Project (Singer, 1972) in its use of military expenditure and personnel data, which have often been used in previous arms race studies. Applying this to cyberspace, the data that is mainly sought here is government spending on cyber security and the number of cyber security specialists employed by governments. These should offer a direct indication of the effort that states are putting into developing their overall cyber strength. Other indicators are relied on if this data is not available. What we aim to indicate is at least whether a significant increase in the effort by states to boost their capabilities is occurring. To determine whether these cyber build-ups are out of line of normal state behaviour, various comparison techniques are used to place them in context.

To answer the second question, a qualitative approach is taken to identify a potential action-reaction and competitive dynamic between our state pairs. We look for a general indication that each state is developing its capabilities in response to the actions of, or the perceived threat posed by, the other. If these criteria are met, it would suggest that there is an arms racing dynamic in cyberspace. While the security portfolio of a state is quite diverse and a major power like the United States likely engages multiple threats, a cyber arms race as we conceptualise it is indicated by the existence of an adversarial relationship and does not demand that all monetary amounts be directed specifically towards the opposition state under examination. The methods we undertake here will allude to the opportunities and challenges in measuring cyber arms races, and our potential limitations are discussed in more depth in the concluding section.

5. THE UNITED STATES AND IRAN

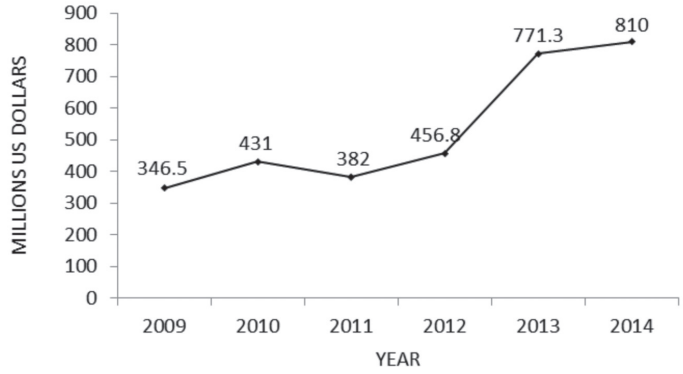
A. US cyber build-up

As a democratic and open society, the United States is relatively forthcoming about its investments in cyber security. The availability of data on two government departments, the Department of Homeland Security (DHS) and the Department of Defense (DOD), allows a rough distinction to be made between the changing defensive and offensive cyber capabilities of the United States.

The DHS is tasked with defending the country against a range of threats, and one of its five stated missions is to 'safeguard and secure cyberspace' by seeking to 'analyse and reduce cyber threats and vulnerabilities […and to…] distribute threat warnings […and…] coordinate the response to cyber incidents to ensure that computers, networks, and cyber systems remain safe' (DHS, 2015). Budget figures are available for the National Cyber Security Division (NCSD), which operates under the Directorate for National Protection and Programs and is home to the United States' Computer Emergency Response Team (CERT) team.

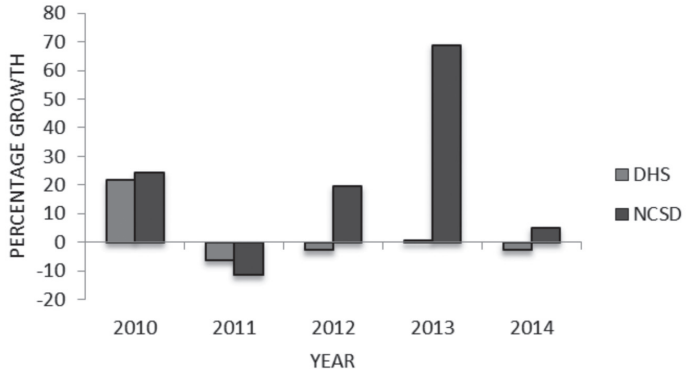
Figure 1 illustrates the changing NCSD budget in constant (2014) US Dollars. Although the unit was formed in 2003, the data is only available between 2009 and 2014.

FIGURE 1: NATIONAL CYBER SECURITY DIVISION BUDGET, 2009-2014 (CONGRESSIONAL RESEARCH SERVICE)



The government funding received by the Cyber Division increased from $346.5 million in 2009 to $810 million in 2014, representing a growth of 134%. The budget has grown in almost every year, with a particularly large jump in 2013. To put these increases in context and determine if it represents an abnormal increase, Figure 2 compares the annual growth in the NCSD budget to that of the DHS as a whole.

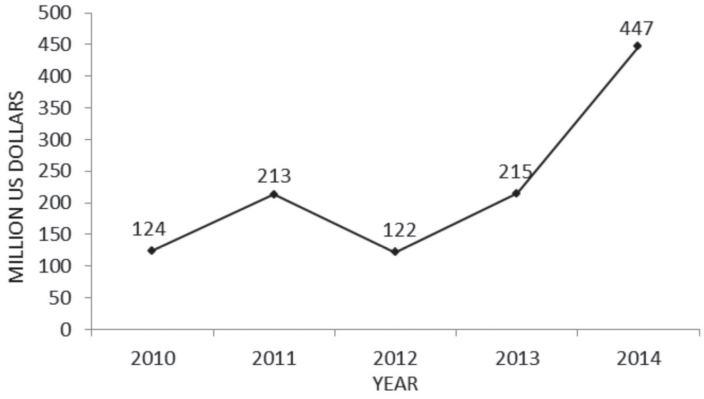
FIGURE 2: ANNUAL GROWTH IN DHS AND NCSD BUDGETS, 2010-2014 (CONGRESSIONAL RESEARCH SERVICE)



On average, the budget of the NCSD grew at higher rates than its parent organisation. The biggest difference came in 2013 when, despite an increase of just 0.4% in the Homeland Security budget, the Division's budget grew by 69% from the previous year.

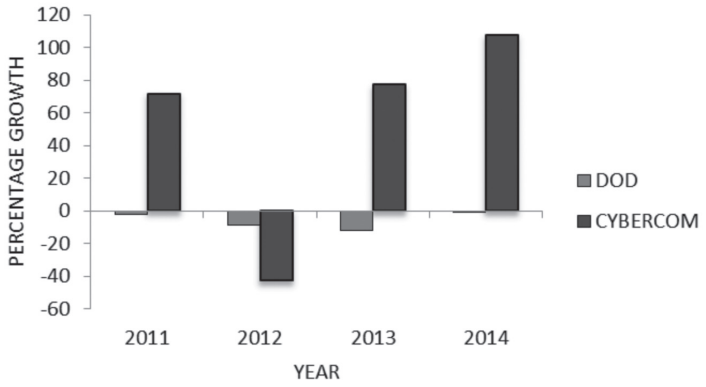
The build-up of offensive cyber capabilities is a more secretive and controversial development, but budget figures are available on the US Cyber Command unit, which reached full operational capacity in 2010. US Cyber Command falls under US Strategic Command, which is one of the 9 military command structures of the DOD. With its stated mission of carrying out the 'full spectrum military cyberspace operations [and to] ensure US/Allied freedom of action in cyberspace and deny the same to our adversaries' (US Stratcom, 2015), the establishment of Cyber Command can be seen as a move to militarise the cyber domain and develop offensive cyber warfare capabilities. Figure 3 shows the changing budget allocation for Cyber Command from 2010 to 2014 in constant US dollars.

FIGURE 3: CYBER COMMAND BUDGET, 2010-2014 (FUNG, 2014)



The US government has evidently been channelling increasing resources into the Cyber Command budget, which has risen from $124 to $447 million since its inception. To give context to this spending pattern, the annual percentage growth in Cyber Command spending is compared in Figure 4 with that of the DOD; in other words, the entire military budget of the United States.

FIGURE 4: ANNUAL GROWTH IN DOD AND CYBER COMMAND BUDGETS, 2011-2014 (FUNG, 2014; SIPRI, 2015)

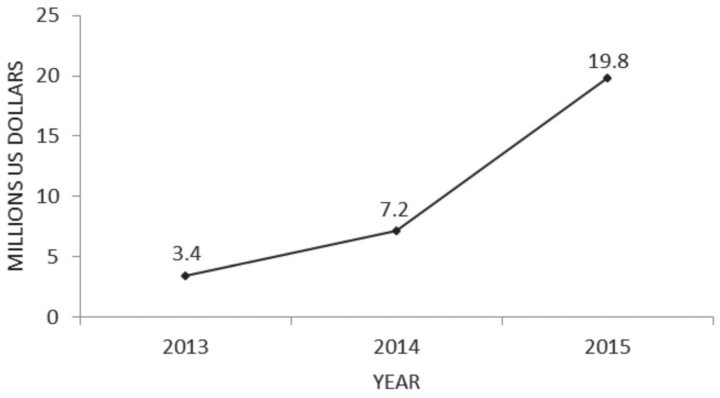


Despite decreases in each year to total defence spending, Cyber Command's budget has tended to grow, and more than doubled in 2014 from the previous year.

B. Iran cyber build-up

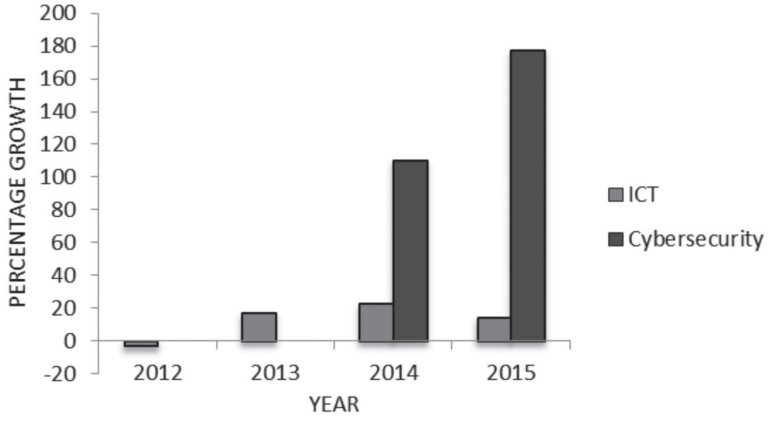
Like the US, Iran is also improving its cyber capabilities. The Iranian Revolutionary Guard Corps has reportedly trained a cyber-army of 120,000 consisting of 'university teachers, students, and clerics', which it claims to be the second largest in the world (UNIDIR, 2013, p.32). In 2012 the Supreme Leader Ayatollah Khamenei established a new cyber unit called the 'Supreme Council of Cyberspace' (SCC), which has ultimate control over all Internet and cyber-related policies in Iran. The SCC's 2014 budget was $40 million, which it receives from Iran's larger ICT budget (Small Media, 2014, p.7). Since President Rouhani came to power, data has been released on Iran's cyber security spending which is presented in figure 5.

FIGURE 5: IRAN'S CYBER SECURITY BUDGET, 2013-2015 (SMALL MEDIA, 2015)



The cyber security budget has increased markedly from $3.4 million in 2013 to $19.8 million in 2015. To put this increase in context, Figure 6 compares the cyber security budget's annual percentage growth in 2014 and 2015 with that of Iran's ICT budget.

FIGURE 6: ANNUAL GROWTH IN IRAN'S ICT AND CYBER SECURITY BUDGETS, 2012-2015 (SMALL MEDIA, 2014; 2015)



Iran's ICT budget has also been increasing year on year since 2013, but not on so great a scale as the cyber security budget. This suggests significant efforts by Iran to specifically improve its cyber capabilities.

C. Dyadic interaction

The United States and Iran have a history of cyber conflict with one another, and as is shown in Table 1, Iran clearly has had more to fear from the United States between 2001 and 2011.

TABLE 1: CYBER CONFLICT BETWEEN THE UNITED STATES AND IRAN, 2001-2011 (VALERIANO AND MANESS, 2014)

US Initiated	6
Iran Initiated	1
Total	7

Initially Iran did not factor much in US cyber strategy, and the sole documented incident carried out by Iran was the 2009 Twitter hack which involved mere website defacement. The competitive cyber relationship was sparked in June 2010 with the discovery of the highly sophisticated Stuxnet computer virus that had been used to target one of Iran's major nuclear enrichment plants in Natanz. The United States, in collaboration with Israel, is widely believed to have masterminded the attack as a means to curb Iran's nuclear ambitions. According to Sanger (2012b, p.205), the attack destroyed 984, or a fifth, of the facility's centrifuges.

Iran's immediate response to Stuxnet was muted, perhaps not wanting to show weakness, yet it soon began developing its cyber capabilities, and in March 2012 Ayatollah Khamenei announced the creation of the Supreme Council of Cyberspace (SCC). Operating under the SCC is the National Centre for Cyberspace (NCC) which is tasked with protecting the country from cyber-attacks, and to help develop a national Internet that would reduce Iran's Internet dependency (Small Media, 2014, p.4).

Retaliation for Stuxnet, and a physical display of Iran's developing offensive cyber capabilities, came in the form of the 'Shamoon' attack, launched by Iran in August 2012 against the Saudi Aramco oil company. Valeriano and Maness (2015, p.157) judge the incident, which deleted data and removed the re-boot program from around 30,000 computers, to be an example of a 'weak state attempting to damage a rival and harm, by proxy, its large state sponsor and greatest consumer of oil'.

After Stuxnet, it became clear that the US feared that Iran was learning from the attack, with the head of Air Force Space Command, General William Shelton, reporting to the media in January 2013 that 'it's clear that the Natanz situation generated a reaction by them'. He identified Iran as 'a force to be reckoned with, with the potential capabilities that they will develop over the years and the potential threat that will represent to the United States'. He also called for increased cyber-security spending, and announced plans to increase the number of cyber personnel in his unit by 1,000 (Shalal-Esa, 2013).

That the US was developing a growing perception of threat from Iran is supported by a Snowdon-leaked NSA document from April 2013. It discussed how Iran had learned from

cyber-attacks launched against it, and had been behind several waves of DDoS attacks on US financial institutions, on top of the Saudi Aramco attack (Greenwald, 2015).

US officials undoubtedly began to see Iran as a source of cyber threat around this time. Speaking before the Senate Intelligence Committee in 2012, the Director of National Intelligence, James Clapper, warned that 'Iran's intelligence operations against the United States, including cyber capabilities, have dramatically increased in recent years in depth and complexity' (Shachtman, 2012). Similarly, in the Committee on Homeland Security in April 2012, it was reported that Iran had invested over $1 billion in expanding its cyber capabilities, and had been carrying out cyber-attacks on media organisations to test its cyber strength (House of Representatives, 2012).

This is a clear example of a state perceiving a threat from the developing capabilities of another, as the action-reaction model predicts. Although a firm connection cannot be proven, it is unsurprising that the data presented on US cyber-warfare spending shows the largest increases after 2012, the year in which the United States apparently became more fearful of the threat from Iran as it responded to Stuxnet. The evidence suggests that both countries developed their capabilities in reaction to one another. Therefore, the competitive aspect of an arms race appears to be present here, as well as the rapid and mutual increase in cyber capabilities.

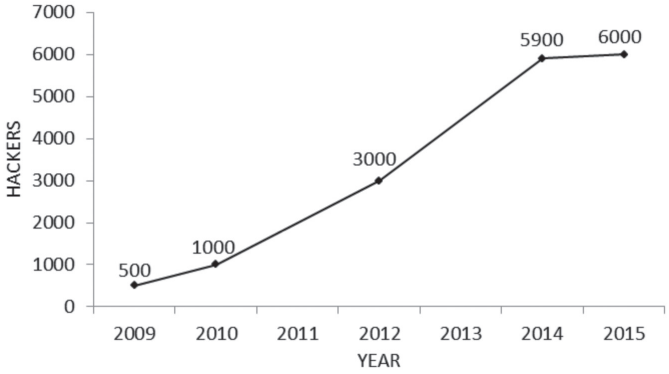
6. NORTH KOREA AND SOUTH KOREA

A. North Korean cyber build-up

The one-party Communist state of North Korea is strongly suspected to be building up its offensive cyber capabilities and is known to have a number of cyber warfare units. Acquiring reliable data on perhaps the most secretive country in the world is particularly challenging. Within the General Staff Department, the Reconnaissance General Bureau runs two main cyber organisations, Unit 91 and Unit 121, both understood to be the source of offensive operations. There are a total of six known cyber units, each with varying cyber warfare roles, including Unit 35 that is believed to be involved in training hackers (Hewlett Packard, 2014, p.26).

A defector to South Korea estimated that between 10 and 20% of North Korea's military budget is spent on 'online operations' (Lee and Kwek, 2015), and a number of defectors as well as South Korean news organisations have made various claims over time regarding the size of North Korea's army of cyber hackers. In figure 7, these estimates are pieced together to highlight the growth of North Korea's offensive capabilities.

FIGURE 7: NORTH KOREA'S 'CYBER ARMY',
2009-2015 (HEWLETT PACKARD, 2014; MULRINE, 2015; LEE AND KWEK, 2015)



If accurate, the data would suggest that the number of North Korean hackers has increased twelvefold since 2009. There is reason to believe such estimates due to the repeated nature of the information, yet the figures are potentially biased since defectors to the south are likely to support heightened concern for North Korean activities.

B. South Korean cyber build-up

South Korea has also been developing its cyber capabilities and in 2010 a cyber-warfare unit was created, staffed by approximately 200 personnel (UNIDIR, 2013, p.41). The data used here to measure South Korea's cyber build-up is the number of secure servers per million of the population. Secure servers are web servers that use encryption technology in Internet transactions, thus somewhat gauging a country's cyber defences. It has certain weaknesses as an indicator of cyber power, but nevertheless appears to show a reaction from South Korea. The change in secure servers from 2003 to 2014 is plotted in Figure 8, and is compared with other groups of countries to put South Korea's cyber build-up into context.

FIGURE 8: SOUTH KOREA'S SECURE SERVERS, 2003-2014 (WORLD BANK, NETCRAFT)

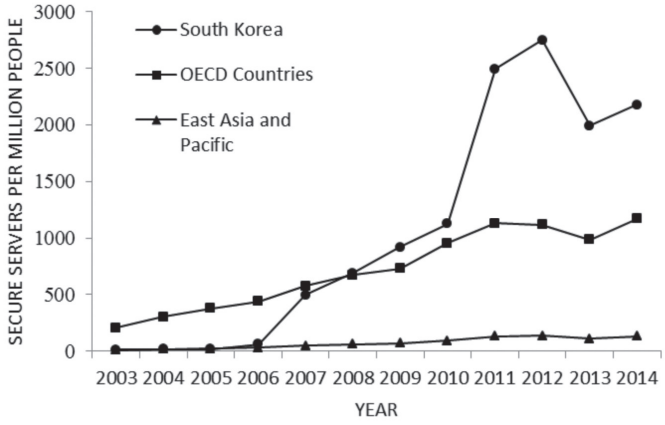


Figure 6 shows a remarkable increase in the number of South Korea's cyber defences. Secure servers grew from just 14 per million people in 2003, to 2,178 per million people in 2014. There was a particularly accelerated period of growth from 2010 when the number of secure servers more than doubled within a year. Furthermore, South Korea's improvements to its cyber capabilities have evidently been on a much greater scale than that of its neighbours in the region, as well as among other economically advanced OECD member states.

C. Dyadic interaction

There have been several known cases of cyber conflict between North and South Korea. Table 2 shows a total of 11 incidents from 2001 to 2011, with North Korea initiating all but one of them.

TABLE 2: CYBER CONFLICT BETWEEN NORTH AND SOUTH KOREA, 2001-2011 (VALERIANO AND MANESS, 2014)

North Korea Initiated	10
South Korea Initiated	1
Total	11

According to the data set, these 10 cyber incidents initiated by the North against the South all took place in the short space of three years between 2008 and 2011, thus giving South Korea a motive for increasing its cyber defences. The process of interaction here has typically been action by the North followed by reaction by the South. For example, in 2009 a DDoS attack hit the networks of several South Korean government organisations and banks (Weaver, 2009). In response, South Korea created a cyber command unit in 2010, with the defence ministry explicitly referencing the threat from North Korea as the justification for the development (Yonhap News Agency, 2010).

South Korea was again targeted by the North in 2011, in an attack that brought down 26 government, military, and banking websites (BBC News, 2011). In the same year South Korea launched its cyber security strategy, now treating the cyber domain as part of the military sphere in the same way as land, sea, or air. Also included in the strategy was a requirement that the public and private sectors take measures to encrypt and back up data (Schweber, 2011). The huge increase in South Korean secure servers from 2010 to 2011 shown in Figure 6 is perhaps directly linked to this policy. In August 2012, the South called for the number of cyber security personnel in its cyber warfare unit to be increased to 1,000 from the 200 initially working there, to help cope with the North Korean threat (Korea JoonGang Daily, 2012).

Another incident in 2013 shut down the South Korean banking system and several television stations. This attack was somewhat more sophisticated in that malware was used, as opposed to the DDoS method, which simply overloads a system with requests (Sang-Hun 2013). This hinted at the growing offensive capabilities of North Korea. In reaction, South Korea announced another build up in manpower, revealing its intention to train an extra 5,000 cyber troops to defend against North Korean cyber-attacks (Hewlett Packard, 2014, p.4). If this was indeed a reaction to the developing capabilities of the North, it gives reason to believe the data

on North Korea's cyber army as it shows South Korea trying to compete with the developments of its rival.

North Korea is by far the more aggressive state in the dyad, but the relationship has not been completely one sided, and the North blamed the South for an attack on its own websites only days before the 2013 attack on South Korea. North Korean State Television referred to the 'intensive and persistent virus attacks [that] are being made every day on Internet servers operated by the DPRK' (Nam, 2013), and warned that they 'will never remain a passive onlooker to the enemies' cyberattacks' (Sang-Hun, 2013).

The presence of a mutual cyber build-up, and the fact that both countries were targeting or responding to one another, is suggestive of an arms racing relationship between North and South Korea also.

7. CONCLUSION AND LIMITATIONS

Both cases appear to meet the criteria for a cyber-arms race which, when applied according to the standards of the international relations research community, is confirmed as a suitable framework for use in the cyber domain. The US-Iran case provides a novel example of cyber competition being driven by mutual insecurity, despite a vast difference in conventional power between them. The actual threat that Stuxnet posed sparked the cyber build-up by Iran, which in turn was perceived as a threat by the United States. The fact that these US security concerns began around the same time as the rapid increases to its cyber security spending suggests they were linked, and that a US-Iran cyber arms race was initiated around 2012. If uncertainty and defensive motivations are indeed at the heart of this cyber arms race, then, given the progress being made on the nuclear issue, there may be hope for an end to its escalation if confidence building measures can be put in place.

The relationship between North and South Korea is somewhat different. Unlike the US-Iran dyad, the insecurity that characterises this arms race has been very one sided since North Korea is motivated in its build-up more by aggressive intent rather than fear. Although there is some indication that North Korea perceived a threat from South Korea, the North is mostly motivated by the desire to cause a nuisance to its long-term rival. This case is an example of an arms race where one state has mainly defensive motives whereas the other has offensive motives, and this creates a difficulty in finding a solution to the escalating competition. In a situation somewhat akin to that of a revisionist power, North Korea is unlikely to give up on its offensive ambitions regardless of levels of threat, which leaves South Korea little choice but to continue to build up its capabilities in response.

This research has demonstrated that there is much the cyber security community can learn from international relations scholarship on arms races. It provides the basis for an understanding of the motivations behind the proliferation of cyber warfare capabilities currently observed in the international system by placing it within the context of interstate competition. Conceptualising this dynamic in cyberspace is an important step in working towards a more secure and

cooperative environment. Given the escalatory nature of arms races, our findings highlight the urgent need for policy makers to understand how their cyber security policies can lead to reactions and create instability as tensions spiral.

Our methodological limitations must be addressed, however. A likely criticism relates to whether we have been able to demonstrate that these cyber build-ups are explained purely by dyadic competition. For instance, surely US cyber spending is motivated just as much, and perhaps more, by its other competitors such as Russia or China. This could very well be true, and we have not tried to argue that its spending is wholly a function of Iranian threat. This is not a necessary criteria for arms races generally, as states must consider all potential threats in the system. Nevertheless, it is clear from our case study evidence that the US perceived a significant threat from Iran and vice versa, which correlates with notable developments in their cyber capabilities.

Since cyber might be an asymmetric domain (Liff, 2012, p.409), we should not dismiss the idea that a traditionally weaker power like Iran plays an important role in US cyber strategy. An extensive report by security firm Cylance even places the Iranian threat on a par with Russia and China (Cylance, 2014). Our analysis leads us to suggest that the term arms race is a reasonable description of the relationship, based on our review of the case. We accept that we cannot demonstrate a causal link between patterns of cyber-build up and the actions or behaviour of another state. To establish such would be difficult without direct statements from the leadership, a condition rare in history. In any study of arms races, it is not possible to calculate just what proportion of spending is accounted for by one particular state, or what threat drives which weapons system. Moreover, factors internal to the state (political, economic, and technological) can have a major effect on military spending patterns, and how this idea relates to the cyber domain is a critical area for future research.

This endeavour is only the beginning of a more systematic investigation of cyber build-ups in international politics. Our method of focusing on single state pairs represents a manageable first step in this particular area, and follows decades of research in the international relations field. At a minimum, we believe we have been able to show that external threats from other states, whether perceived or real, are an important variable in shaping a state's national cyber security policies. We aim to build on what we have begun here and continue to identify the wider range of factors accounting for the acquisition of cyber capabilities.

The next step will include expanding the number of cases, collecting data on a wider range of indicators, and developing a methodology for accurately judging cyber power. Despite the secrecy that pervades the domain, the collection and analysis of data relating to cyber security is possible, although difficult and time consuming. It is nevertheless a much needed task if we are to ground the study of cyber conflict within empirical research frameworks.

REFERENCES

Bumiller, Elizabeth, and Thom Shanker, 'Panetta Warns of Dire Threat of Cyberattack on U.S.', The New York Times, 11 October 2012, http://www.nytimes.com/2012/10/12/world/panetta-warns-of-dire-threat-of-cyberattack.html

Choucri, Nazli. 2012. Cyber Politics in International Relations, (Cambridge: MIT Press)

Clarke, Richard A. and Robert K. Knake. 2010. Cyber War: The Next Threat to National Security and What to Do About It, (New York: Harper Collins)

Congressional Research Service Reports on Homeland Security, last updated December 15, 2015, http://www.fas.org/sgp/crs/homesec/

Corera, Gordon, 'Rapid escalation of the cyber-arms race', BBC News, 29 April 2015, http://www.bbc.co.uk/news/uk-32493516

Cylance, 'Operation Cleaver', December 2014, https://cdn2.hubspot.net/hubfs/270968/assets/Cleaver/Cylance_Operation_Cleaver_Report.pdf

Eckstein, Harry. 1975. 'Case Study and Theory in Political Science', in The Handbook of Political Science, eds. F. I. Greenstein and N. W. Polsby, (Reading: Addison-Wesley)

Fung, Brian,'Cyber Command's exploding budget in 1 chart', The Washington Post, 15 January 2014, https://www.washingtonpost.com/news/the-switch/wp/2014/01/15/cyber-commands-exploding-budget-in-1-chart/

George, Alexander L. and Andrew Bennett. 2005. Case Studies and Theory Development in the Social Sciences, (Cambridge: MIT Press)

Gibler, Doug, Toby J. Rider, and Michael Hutchison. 2005. 'Taking Arms Against a Sea of Troubles: Conventional Arms Races During Periods of Rivalry', Journal of Peace Research, 24(2): 251-276

Glaser, Charles L. 2000. 'The Causes and Consequences of Arms Races', Annual Review of Political Science, 3: 251-276

Glaser, Charles L. and Chaim Kaufmann. 1998. 'What is the Offense-Defense Balance and Can we Measure it?', International Security, 22(4): 44-82

Gray, Colin S. 1971a. 'The Arms Race Phenomenon', World Politics, 24(1): 39-79

Greenwald, Glen, 'NSA Claims Iran Learned from Western Cyberattacks', The Intercept, 10 February 2015, https://firstlook.org/theintercept/2015/02/10/nsa-iran-developing-sophisticated-cyber-attacks-learning-attacks/

Hammond, Grant T. 1992. Plowshares into Swords: Arms Races in International Politics, 1840-1991, (Columbia: South Carolina Press)

Hewlett Packard, 'Profiling an enigma: The mystery of North Korea's cyber threat landscape', HP Security Briefing Episode 16, August 2014, http://h30499.www3.hp.com/hpeb/attachments/hpeb/off-by-on-software-security-blog/388/2/HPSR%20SecurityBriefing_Episode16_NorthKorea.pdf

Horn, Michael Dean, 1987. 'Arms Races and the International System', PhD diss., (Rochester, NY: Department of Political Science, University of Rochester)

House of Representatives, 'Iranian Cyber Threat to the U.S Homeland', Joint Hearing before the Subcommittee on Counterterrorism and Intelligence and the Subcommittee on Cybersecurity, Infrastructure Protection, and Security Technologies, 26 April 2012

Huntington, Samuel P. 1958. 'Arms Races: Prerequisites and Results', Public Policy, 8: 1-87

Jervis, Robert. 1976. Perception and Misperception in International Politics, (Princeton, Princeton University Press)

Jervis, Robert. 1978. 'Cooperation Under the Security Dilemma', World Politics, 30(2): 167-214

Kello, Lucas. 2013. 'The Meaning of the Cyber Revolution: Perils to Theory and Statecraft', International Security, 38(2): 7-40

Lee, Dave, and Nick Kwek, 'North Korean hackers 'could kill', warns key defector', BBC News, 29 May 2015, http://www.bbc.co.uk/news/technology-32925495

Leeds, Brett A. and T. Clifton Morgan. 2012. 'The Quest for Security: Alliances and Arms', in Guide to the Scientific Study of International Processes, Ed. Sarah McLaughlin Mitchell, Paul F. Diehl, and James D. Morrow, (Wiley-Blackwell)

Liff, Adam P. 2012. 'Cyberwar: A New 'Absolute Weapon'? The Proliferation of Cyberwarfare Capabilities and Interstate War', Journal of Strategic Studies, 35(3): 401-428

Lindsay, Jon R. 2013. 'Stuxnet and the Limits of Cyber Warfare', Security Studies, 22(3): 365-404

Mulrine, Anna, 'How North Korea built up a cadre of code warriors prepared for cyberwar', Christian Science Monitor, 6 February 2015, http://www.csmonitor.com/World/Passcode/2015/0206/How-North-Korea-built-up-a-cadre-of-code-warriors-prepared-for-cyberwar

Nam, In-Soo, 'North Korea Complains of Cyberattacks', The Wall Street Journal, 15 March 2013, http://blogs.wsj.com/korearealtime/2013/03/15/north-korea-complains-of-cyberattacks

Nye, Joseph. 2011. The Future of Power, (New York: Public Affairs)

Richardson, Lewis F. 1960. Arms and Insecurity: A Mathematical Study of the Causes and Origins of War, ed. Nicolas Rashevsky and Ernesto Trucco, (Pittsburgh: The Boxwood Press)

Rid, Thomas, and Peter McBurney. 2012. 'Cyber-Weapons', The RUSI Journal, 157(1): 6-13

Sample, Susan. 1997. 'Arms Races and Dispute Escalation: Resolving the Debate', Journal of Peace Research, 34(1): 7-22

Sanger, David E. 2012b. Confront and Conceal: Obama's Secret Wars and Surprising Use of American Power (New York: Random House)

Sang-Hun, Choe, 'Computer Networks in South Korea Are Paralyzed in Cyberattacks', The New York Times, 20 March 2013, http://www.nytimes.com/2013/03/21/world/asia/south-korea-computer-network-crashes.html

Schweber, Aerianna, 'South Korea Develops National Cyber Security Strategy', Intelligence, 28 August 2011, http://blogs.absolute.com/blog/south-korea-develops-cyber-security-strategy

Shachtman, Noah, 'Iran Now a 'Top Threat' to US Networks, Spy Chief Claims', Wired, 31 January 2012, http://www.wired.com/2012/01/iran-now-a-top-threat-to-u-s-networks-spy-chief-says/

Shalal-Esa, Andrea, 'Iran strengthened cyber capabilities after Stuxnet: US General', Reuters, 17 January 2013, http://www.reuters.com/article/2013/01/18/us-iran-usa-cyber-idUSBRE90G1C420130118

Singer, J. David. 1972. 'The 'Correlates of War' Project: Interim Report and Rationale', World Politics, 24(2): 243-270

Small Media, Iranian Internet and Infrastructure Policy Report, February 2014, http://smallmedia.org.uk/sites/default/files/u8/IIIP_Feb2014.pdf

Small Media, Iranian Internet Infrastructure and Policy Report, January 2015, http://smallmedia.org.uk/sites/default/files/u8/200215_InternetInfrastructure%20(1).pdf

SIPRI, Military Expenditure Database, last updated November 2015, http://www.sipri.org/research/armaments/milex/milex_database

Thompson, William R. 2001. 'Identifying Rivalry and Rivalries in International Politics', International Studies Quarterly, 45(4): 557-586

UNIDIR, The Cyber Index: International Security Trends and Realities, http://www.unidir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf

Valeriano, Brandon, and Ryan C. Maness. 2014. 'The dynamics of cyber conflict between rival antagonists, 2001-11', Journal of Peace Research

Valeriano, Brandon, and Ryan C. Maness. 2015. Cyber War versus Cyber Realities, (New York: Oxford University Press)

Valeriano, Brandon, Susan Sample, Choong-Nam Kang. 2013. 'Conceptualising and Measuring Rapid Military Buildups in the International System', Presented at Eurasian Peace Science Conference, Istanbul, Turkey, May 24-25 2013

World Bank/Netcraft, Secure Internet Servers (per 1 million people), last accessed December 2015, http://data.worldbank.org/indicator/IT.NET.SECR.P6

Weaver, Matthew, 'Cyber attackers target South Korea and US', The Guardian, 8 July 2009, http://www.theguardian.com/world/2009/jul/08/south-korea-cyber-attack

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Arms Control in Cyberspace – Architecture for a Trust-Based Implementation Framework Based on Conventional Arms Control Methods

Markus Maybaum
Fraunhofer FKIE
Bonn, Germany
NATO CCD COE
Tallinn, Estonia
markus.maybaum@fkie.fraunhofer.de

Jens Tölle
Fraunhofer FKIE
Bonn, Germany
jens.toelle@fkie.fraunhofer.de

Abstract: This paper explores verification mechanisms, as well as confidence and security building measures, within the scope of existing conventional and strategic arms control regimes. In particular, it analyses the concepts of the Conventional Forces in Europe Treaty and the Vienna Document as an implementation regime for confidence and security building measures as well as the Open Skies Treaty, representing three major conventional arms control regimes. As an example for strategic arms control, we analyse the Chemical Weapons Convention. The objective of this paper is to identify those means and methods from these successful frameworks that can be adapted and potentially incorporated into a cyber-domain arms control regime. Based on this discussion, the authors suggest a general technical architecture for a trust-based future framework for arms control for cyberspace.

Keywords: *arms control, cyberspace, trusted computing, advanced trusted environment*

1. INTRODUCTION

Arms Control (AC) has been a success story since the late 1980s. While the first multilateral AC regimes focused more on conventional weapons and confidence-building measures in the air, space, and sea domains, recent treaties have been grander in scope. Consequently, arms control in cyberspace (ACCS) is seen as a necessary next step in building confidence between arising cyber powers operating within the cyber domain. AC is a commonly recognised instrument of security policy to avoid a competitive build-up of weapons between powers – an arms race [1]. Such a race is always costly for all sides. AC treaties have been negotiated and set into force to serve the purpose of limiting weapons stockpiles to a level that promises deterrence while conserving the economic and social resources of a state for other uses. When, more than 40 years ago, the Helsinki Final Act was signed, no one could really foresee the positive de-escalation between the main parties of the Cold War: NATO and the Warsaw Pact.

The story of success can be seen in more than three dozen AC treaties that have since been negotiated. However, recent developments in global security policy indicate a substantial change. Traditional AC has become subject to criticism and, especially from an American perspective, conventional AC in Europe has been disparaged [2]. This is not only due to the changed security situation in today's Europe but it is also an outcome of the inability of the international community to further advance these important security-enhancing instruments. The practice of hybrid warfare [3] has been underlying conventional AC regimes long before the Ukraine crisis – without any further action being taken or there being a need to implement suitable security mechanisms as a reaction to those demonstrated scenarios. Threats arising from cyber campaigns are no longer science fiction, and cyber weapons are undoubtedly a new instrument in conflict scenarios.

A. Aim

Besides obvious political obstacles, the development of an ACCS regime needs to cope with substantial practical and technical challenges. From the technical perspective, both security and privacy will be the most significant issues that must be addressed. On the practical side, the key question is enforcement: how can states make sure that any provisions are implemented? [4] Being aware of the complexity a regime to effectively monitor treaty compliance would require [5], an overwhelming majority of researchers in that field have all but given up; some have even concluded that ACCS in practice will be almost impossible [6].

This paper does not share that opinion. We think that, especially within the scope of building future cyberspace, there are several options to plan and implement instruments for ACCS. Conventional AC was seen as unlikely in the early 1980s, but less than a decade later the first international agreements on AC and verification had been put in place. Consequently, we see ACCS not only as a possibility, but also as a necessary next step in building confidence between rising cyber powers. It is also apparent that global players are starting to share our views to a certain extent. For example, we have seen international agreements such as the Wassenaar Arrangement [7] prohibiting the export of dual-use goods and technology, including computers and means of information security. A number of bilateral agreements have also been signed

between different nations. The first impulse towards an ACCS regime came from Russia, and their effort is still on-going [8]. In 2015 China and the USA negotiated a political agreement committing each country not to be the first to use cyber weapons to attack the other's critical infrastructure during peacetime [9], following the spirit of the Helsinki Final Act.

This paper's objective is to identify those means and methods that can be adapted from the successful AC frameworks and potentially incorporated into an ACCS regime. It recognises the functional gap between the identified and implemented requirements and evaluates the implications that arise from this difference analysis of future ACCS. As a proof of concept, a general technical architecture for a trust-based future framework for ACCS is suggested.

B. Definitions

Before we start analysing the AC regimes, we need to specify the key terms used in this paper. We will mostly refer to existing definitions that have been accepted in the academic community with some specific adaptations that may be required. For the purpose of this paper:

(a) A 'Weapon' is 'a means of warfare used in combat operations, including a gun, missile, bomb or other munitions, that is capable of causing either (i) injury to, or death of, persons; or (ii) damage to, or destruction of, objects' [10].
(b) A 'Weapon of Mass Destruction' is a nuclear weapon, a biological weapon or a chemical weapon as defined in [11].
(c) A 'Conventional Weapon' is a Weapon that is not a Weapon of Mass Destruction.
(d) A 'Cyber Weapon' (CyW) is a Weapon that would comprise any computer equipment or device that is designed, intended or used to have violent consequences; that is, to cause death or injury to persons or damage or destruction of objects [12].
(e) A 'Cyber Arm' as a CyW that is used in a computer network attack.

2. ANALYSING TRADITIONAL ARMS CONTROL REGIMES VS. CYBER ARMS CONTROL

In this chapter, we analyse well-known and established AC regimes. In particular, we identify ideas, methods, and techniques that have been deemed successful. We will explain the core objectives of four active AC regimes and highlight the main parameters based on which the treaty was designed and implemented. We will summarise these objectives and demonstrate the analogies to cyberspace.

A. The Conventional Forces in Europe Treaty and the Vienna Document

1) Main objectives

The Conventional Forces in Europe (CFE) Treaty, signed in November 1990, outlines provisions aimed at establishing a military balance between NATO and the Warsaw Pact, at a lower level of armaments [13]. It was negotiated during the late 1980s when NATO and the Warsaw Pact were both focusing on ending the arms race between East and West. The Vienna

Document (VD) is another regime of AC and confidence-building negotiated between the 57 participating states of the Organisation for Security and Co-operation in Europe (OSCE) [14]. It was adopted in 1990, together with the drafting of the CFE Treaty, and underwent its most recent fundamental revision in 2011 [15].

Both CFE and VD focus on conventional weapons in Europe. With CFE, the term Treaty Limited Equipment (TLE) was introduced with a broad scope, defining exact categories of weapons systems for which the treaty was supposed to be applicable. One of the typical characteristics of the CFE and other conventional AC treaties is that the main focus was laid on the weapons systems themselves and not, for example, on their ammunition. As we will see later in the context of the Chemical Weapons Convention (CWC), ammunition is only discussed separately when its damage potential is very high, such as for landmines or weapons of mass destruction. Based on the defined TLE, the CFE Treaty foresees an initial declaration of all TLE owned by a nation, its home location, and the military unit it is assigned to, including a full layout of force structure, followed by a yearly update. The establishment of a military balance between NATO and the Warsaw Pact was achieved by structural and geographical provisions such as defining limits for specific types of TLE, not only in total but also in certain concentric zones bearing in mind that for a successful attack TLE would have to be moved. Thus, the CFE regime in particular has an early warning function. To meet these limitations, equipment had to be destroyed or converted to non-military purposes.

The VD offers transparency and confidence-building by a declaration and inspection regime as well as mechanisms for peaceful conflict resolution. In signing the VD, the OSCE member states committed themselves to submit detailed information on their armed forces and principal weapons systems, their military budgets, their defence and force planning, and military exercises to the other state parties on an annual basis. OSCE states can conduct confidence-building inspections to verify this submitted information for compliance with the provisions of the VD. In addition to on-site inspection agreed in the CFE Treaty, VD allows so called ‘inspections of specified areas’ within the territory of the inspected state. This mechanism allows tracking of military activities that are taking place in these areas. The inspection team is entitled to check such an area on the ground and from the air. In addition, the VD foresees the invitation of observers to manoeuvres once they exceed a defined size, and it requires its member states to announce and present new weapons systems that they introduce.

The confidence-building function of both treaties was based on two main pillars: verification of both the declaration and yearly updates by on-site inspections, and social interaction between the inspection teams during their routine work. The CFE regime also established a so-called consultation group that had to deal with treaty interpretations and complaints. The consultation group was the official communication platform between the member nations. With Russia’s suspension of the CFE Treaty in 2007 [16] and subsequent loss of transparency around conventional forces, the politically binding procedures and related reports associated with the VD have become more important [17]. The signing parties agreed to hold an Annual Implementation Assessment Meeting where states are given a platform to discuss questions of implementation, operations, and questions that may have arisen from information that has

been exchanged or from findings during inspections. This meeting is hosted by the Forum for Security Cooperation [18] and has especially been used to discuss the situation in Ukraine.

2) Applying CFE and VD techniques in cyberspace

When thinking about applying CFE and VD techniques in cyberspace, the first issue that needs to be discussed is the definition of TLE and their categorisation. The core functions and principles of CFE refer to conventional weapons systems, counting these systems, and declaring possession and location of them. Considering our definition of a CyW and also other common definitions (e.g. 'software, firmware or hardware designed or applied to cause damage through the cyber domain' [19]), these core functions would only be applicable if hardware is involved; CyWs consisting of software are obviously uncountable [20].

The possession of such CyWs can be declared, of course, but any structural or geographical provisions or limits would not make sense or could not be subject to any form of on-site inspection. Thus, the early warning function, similar to the CFE does would not work efficiently in cyberspace and must be seen as a functional gap. The same applies to verification and social interaction; since any on-site inspection regime in combination with a declaration regime does not make sense within the scope of cyberspace, this function must also be seen as quite limited. The role of a consultation group within the scope of ACCS would be different from the role of a CFE consultation group; we will elaborate on this in more detail in section 3.

Nevertheless, the idea to allow areas to be subject to inspection can be of interest when thinking of inspecting parts of a state's cyberspace; for example, autonomous systems or parts of the national network infrastructure. Invitations to manoeuvres may also be an option when considering exercises or manoeuvres taking place in cyberspace, with the option of having observers present (physically or virtually). The challenge here is the level of detail an observer would be invited to see: would a cyber inspection team be granted read-access to inspection networks? How can espionage be prevented in such a setup? This raises essential questions that need to be answered.

B. Open Skies Treaty

1) Main objectives

The Open Skies (OS) Treaty established a regime of unarmed observation flights over the territories of states [21]. The idea of an airborne verification regime was born during the Cold War, but it never left the blueprint-stage due to mistrust within NATO and the Warsaw Pact; both sides were afraid of potential espionage. The OS Treaty was signed in March 1992. The ratification process in the Russian Federation took 10 years, as many technical details had to be discussed, tested and agreed on due to concerns over espionage. Finally, the treaty was put in force on January 1st, 2002.

The OS Treaty defines observation flights that an inspecting state party can conduct over the territory of another state party. The inspected state must provide airfields for that purpose from which those observation flights can be launched. One of the major objectives of OS is

the territorial scope: the observation flights can cover the entire country – thus, except from force majeure or natural conditions that would make flights impossible, no areas of a state party's territory can be excluded. An observation flight has to be notified three days in advance, specifying a point of entry (a predefined OS airfield). The detailed route is not submitted in advance, but negotiated with the inspected state party after arrival.

Besides quotas for observation flights and the notification of points of entry, the technical details and inspection of sensors were seen as critical during the negotiation of the OS Treaty. All parties involved were interested that the sensors of the observation aircraft could only be used for their dedicated purpose. This implies that all sensors must have specifically defined and technically assured parameters such as a maximum resolution which makes it possible to identify TLE, but not to record details of intelligence value. For this, all sensor configurations of an observation aircraft have to be certified, which in terms of the OS Treaty means that they have to be validated against a calibration target that allows an exact calculation of the camera resolution before an observation aircraft is permitted to conduct observation missions. The aircraft is also subject to inspection before missions to that ensure no additional sensors are hidden on board.

The confidence-building function for the OS Treaty is, by design, based mostly on social interaction between the mission teams during their routine work as well as the specialists working together during the specification of the technical parameters, calibration, and testing. Since the OS Treaty has no own declaration regime, it works more as a service for other verification regimes, supporting their confidence-building functions by providing the results of the observation missions to the teams. As an additional confidence-building measure, the results of the observation missions are provided to all OS Treaty participants.

2) Applying the OS Treaty's techniques in cyberspace

Referring to the main principles of the OS Treaty, some useful lessons can be learned for a possible ACCS regime. The first characteristic is the territorial scope: the entire country is subject to inspection. In the context of cyberspace, this would mean that the entire cyber infrastructure of a country would be subject to inspection. Still, software can be easily shipped and stored outside the country's borders, so there is still a functional gap as long as there are states not participating in such an ACCS regime. The same applies to small mass production hardware components such as microprocessors or other microchips. Malicious code can be stored and hidden easily, unless such a hidden functionality can reliably be found by technical means.

Serious challenges are posed if we take into account the technical details and certified sensors needed for ACCS mechanisms. Simply the overwhelming amount of known existing malware and the exponential increase of new examples may serve as an indicator that the traditional ways of finding malicious software by technical means have reached their limits. The functional gap we see here is a technical solution to reliably identify a CyW in cyberspace. For this, a new technology is required which experts can negotiate in order to find suitable technical parameters which can serve as the technical core of a future AC regime for cyberspace. Many

experts see this as the main argument against ACCS since finding reliable metrics for CyW detection is considered to be impossible, and large anti-virus companies have already given up this arms race [22]. So, the lesson we can learn from OS is that it could be used as an example to develop a cooperative approach of world leading experts working together to solve the malware problem. This would also surely support the confidence-building function we have seen in OS, gaining trust at the expert level by working together and developing a common platform that can make ACCS possible.

C. The Chemical Weapons Convention

1) Main objectives

The CWC is a strategic AC regime with a global scope [23] that was signed in 1993 and came into effect in 1997. It aims to eliminate all existing chemical weapons (CW) globally by prohibiting the development, production, acquisition, stockpiling, retention, transfer, or use of CWs by state parties. The first challenge with CWs is the exact definition of the TLE: what is a 'chemical weapon'?

The CWC's approach is to define lists of agents that fulfil certain requirements. In particular, typical agents for the use within the scope of chemical warfare are listed in relation to the Schedule 1 category. For the classification of CWs, the CWC introduced 3 categories [24]:

(a) Category 1: CWs based on Schedule 1 chemicals, including VX and Sarin (See below for an explanation of 'scheduled' chemicals);
(b) Category 2: CWs based on non-Schedule 1 chemicals, such as phosgene;
(c) Category 3: CWs including unfilled munitions, devices, and equipment designed specifically to employ CWs.

According to the CWC, a member state must declare its possession of CWs in an initial declaration. State parties are then obliged to plan and organise the destruction of their CWs. Since the destruction needs specific facilities to be built, and budget and administrative issues have to be solved, the joining state party negotiates an action plan with the Organisation for the Prohibition of Chemical Weapons (OPCW). The OPCW is an international implementation agency and has the role of a convention management entity supervising treaty implementation. In this action plan, a detailed time line with milestones is defined explaining how a member intends to eliminate its declared CWs. The progress of CWs' destruction is again subject to a notification regime, and the CWs, their storage facility, and the destruction facilities are subject to regular inspection.

Additionally, the CWC foresees so-called challenge inspections. If a CWC member accuses another member of false reporting of its CWs arsenal or any details of the negotiated action plan, the OPCW can initiate an area- or on-site inspection that the accused has to accept. Besides these technical specifications around a traditional declaration and inspection regime that have an obvious similarity to conventional AC systems, CWC has a global scope: so far, only four states (Egypt, Israel, North Korea, and South Sudan [25]) have not ratified the treaty.

This shows remarkable political will to do away with this class of weapon of mass destruction.

2) Applying CWC techniques in cyberspace

The core principle of CWC is not to focus on describing the CWs themselves, but to describe the agents these weapons carry as a payload. The three categories introduced in the CWC define a priority list. It is based on the severity of impact a weapon could have which is determined by the specific payload of the weapon.

CyWs can be seen as similar: the detailed malicious function may be undetectable until the weapon is launched, but the impact of a CyW in a specific attack scenario can be described in pre-defined metrics [26]. In general, the impact of a CyW will be either a breach of integrity of a system, or limitation to its availability; this is what an ACCS regime will have to detect.

The political will is also a major objective when thinking about regulating cyberspace [20]: will a majority of states be willing to share the idea of a peaceful use of the cyber domain? Will they be willing to invest in a common supervising entity such as the OPCW governing the future development of secure cyberspace and prohibiting warfare within the net? An action plan for weapon destruction will not be required if the common goal of nations is the peaceful cooperative use of modern information infrastructure.

The development of reliable technologies capable of monitoring malicious activities will be the key to success, so the question arising from the idea of the CWC relates back to a core technical problem: will we be able to develop a technical framework that allows us to identify CyWs in cyberspace? What would be the equivalent of a CWC challenge inspection? Would it be a reliable procedure based on globally trusted digital forensics that can prove or disprove a CyW engagement? We will address these questions in the following section.

3. CONCEPTS FOR THE IMPLEMENTATION OF A CYBER ARMS REGIME BASED ON TRADITIONAL ARMS CONTROL METHODS

When analysing the ideas and techniques of traditional AC in the context of cyberspace, we found potential analogies but also functional gaps that we see as requirements for a future ACCS regime. We discuss the analogies and gaps in the next section. Based on these findings, we then introduce a technical framework that we believe is helpful to make ACCS possible.

A. Analogies and gaps of traditional arms control vs. arms control in cyberspace

The core function of the first conventional AC treaties was the establishment of an early warning function, realising that the preparation of an armed attack in preparation of a future war or armed conflict would require movement and assembling of significant parts of a state's armed forces. In cyberspace, the situation is different.

CyWs cannot be clearly identified by an inspection team or a technical sensor due to their

characteristics. In the worst case, a CyW consists of numerous distributed pieces of information that are assembled at the moment of the attack. Thus, an approach to finding an early warning mechanism does not appear promising. Since, at the same time, a pre-planned cyber operation can be launched almost instantaneously, an ACCS regime would need a real-time warning mechanism. What makes it even worse is that without knowing potential patterns of CyWs, any information could be a potential suspect. Looking for 'dangerous parts' of software, similar to the list of dangerous agents within the CWC, is like searching for a needle in a hay stack, especially if encryption needs to be considered. Thus CyWs can only be detected if they have been seen before, or simply by coincidence. The dangerous weapons – based on the principle of unknown vulnerability exploitation – can mostly not be found this way.

On the other hand, hope cannot be a plan, meaning that for any further thought about an AC regime at the technical level, the focus at the current state of technology should not be on detecting the CyW itself but on its engagement or effect. In this respect, inspection of sites or areas, as we have seen in the traditional AC regimes, are unrealistic. The earliest possible time for a reliable detection of a CyW is when it is fully assembled and has a payload. This requires AC taking place at the 'speed of cyber'; we call this ad-hoc arms control. Technically speaking, we need to identify the breach of system integrity or the crippling of its availability. This can be achieved by technical means, as we will demonstrate in section 4.

The need to discover such integrity breaches or availability degrades already implies the use of sensors, as we have seen CWC and especially in OS before. Within the OPCW, experts of all nations share their knowledge on dangerous agents as well as on technologies of detection and CW disposal. The OS Treaty regime became successful precisely because technical experts of all parties worked together, and had the same goal of making it possible. This was achieved by openness, technical concepts, and a fully transparent framework. This also requires a technology that fulfils the aims and scope of AC which would not allow back doors or unauthorised use at the same time, as is the case with cyber espionage. In the context of cyberspace, we see the challenges arising from the development and implementation of such a framework as being even more complex. We think that an ACCS regime can be possible, if the political will is there and if persuading ideas for such a framework can be adequately promoted. One of the main challenges within this process will be the involvement of a majority of states around the globe, since geographical limitations within the scope of cyberspace are meaningless.

B. A common interest of states for stability and peace in cyberspace?

We can learn from the CWC that if mankind recognises an imminent threat beyond national borders, a common policy may be achieved. The objective of ACCS is to limit or stop a cyber arms race and to permit the peaceful use of cyberspace, which should be in the common interest of everybody. This is rather easy to demonstrate simply when showing people what a step back from the concepts of a digital information society to an analogue world would entail. Taking away the Internet from mankind would have a global impact on civilisation with unforeseeable consequences [27], providing reason enough to preserve this new territory and to establish common international security concepts and policies, including AC as an established concept of success. We see national legal frameworks and regulations, and authorities being assigned

responsibilities for national parts of the cyber domain. However, we do not see its equivalent for consultation boards as we have seen them for the traditional arms regimes. First discussions on confidence-building measures have been made at OSCE level [28], but so far we do not see groups of policy and technical experts cooperating to work on concepts for stability and peace in cyberspace. Cooperation between policy and technical experts is of major importance. The OS Treaty showed that establishing trust is the key, and that cooperation during the development of the technical framework was partly more effective in achieving confidence-building than the actual implementation of the jointly developed concepts.

By its nature, trust in cyberspace has a very technical dimension. The key is the development of reliable technologies. These technologies need to be transparent, so that their functionality can be understood by all parties involved. The technologies must guarantee the demands for confidentiality of states as well as their citizens and entities. The main challenge and obstacles with the Biological Warfare Convention (BWC) [29], for example, addressed exactly this point: the expert groups negotiating the implementation details were not able to find suitable technologies and procedures to establish an effective AC regime due to the difficulties with guaranteeing confidentiality of the business and state secrets of the inspected party. As a result, the BWC negotiations about the establishment of a binding verification regime have never succeeded and hopes for a global and verifiable prohibition of biological weapons, similar to the CWC, have unfortunately been abandoned. This will be another challenge for an ACCS regime: finding a technical solution to identify integrity breaches without getting to know too many details of the breached system. We will discuss and propose a possible solution addressing this concern in the next section.

Besides trust in technology, trust in science is a mandatory requirement when thinking about ACCS. Whereas in traditional arms control regimes the procedures and technologies could be demonstrated and made understandable to a broad audience, these concepts in cyberspace are significantly more abstract and a closer understanding cannot be expected either by political leaders or military decision makers. If the design of a future ACCS regime is too technically sophisticated, these decision making levels need to rely on experts they may or may not have. Taking this psychological aspect into account, we also see the requirement to design the technical framework based on technical standards that are internationally recognised and accepted. In our suggestions for a technical framework for future ACCS, we therefore refer to commonly developed international standards and enhance these standards to fulfil the necessary requirements.

4. A FUTURE ADVANCED TRUST ENVIRONMENT FOR ARMS CONTROL

Simply put, the core problem with a possible ACCS is to find the needle in the haystack without knowing what the needle looks like. In traditional AC, we know the needle, and in strategic regimes such as BWC and CWC we know at least the shape of the needle. Therefore, we can determine if certain identifiable parts can be part of such a needle. In the cyber domain, we do not know anything but the fact that we look for something having the function of a needle.

CyWs – like any other code – consists of many small 'puzzle' pieces of code. No one can say for sure if a piece of code is malicious without seeing the entire picture – or at least a big part of it.

All AC regimes with verification mechanisms we have seen so far work with a 'black list' describing the limited or prohibited item they control. This will not work with the infinite amount of potential malicious code pieces that can be merged to an uncountable number of different samples. In order to find working solutions, we have to reverse our thinking. A successful technical implementation framework for future ACCS will thus have to adopt a 'white list' approach that focuses on identifying good code from malicious code. This requires trust relationships in the levels of human-machine as well as machine-machine that have not yet been implemented in a broad scope.

As an enabler for future ACCS, we therefore suggest the development of a technical framework which we call Future Advanced Trusted Environment (FATE) to implement the requirements we defined in section 3.A based on technical white-listing. As we have already concluded, a peaceful use of cyberspace can be made possible by the early detection and prevention of the use of CyWs. Taking this into account, the application of CyWs in our context has to be technically understood as the execution of malicious code that breaks the integrity of the affected system or at least degrades its availability. Thus, we are here focusing on the challenges arising from this task.

To deliver FATE, both the computers as well as the communication links between the computers have to establish a Trusted Network (TN) that would technically implement such white-listing. The general idea behind this concept is that FATE must enable the real-time detection of any integrity breach during system operations as well as monitor and report any limitations on availability. Probing of availability and deriving a Common Availability Pictures (CAP) is daily practice in Computer Emergency Response Teams and Network Operation Centres around the globe. Monitoring integrity breaches at the tactical level and compiling a Common Integrity Picture (CIP) to form a Common Cyber Situation Picture (CCSP) is a functional gap that needs to be solved.

If the CIP is effectively established, good code can then be distinguished from malicious code (CyWs) even without knowing the details of the software products being used or the data being processed. This can be achieved by obtaining technical metadata that can be used to annotate programs and data. In current standard architectures, any code is processed. Malware is brought in as data and it hijacks a regular control flow by deviating the processing of the application into malicious code. Our approach is to counter the software exploitation exactly at that point: by defining which branches of an application control flow are legitimate during execution (white-listing), we can prohibit any control-flow deviation into malware – no matter what system is being used and what software is being executed.

Being aware of the need for such a white-listing-based architecture, we analysed the established concept of Trusted Computing (TC) [30] as the state-of-the-art hardware architecture in that field. In TC, all installed software (including BIOS and drivers) is digitally signed and any

unauthorised modifications can be recognised. Based on its hardware extension, the Trusted Platform Module (TPM), we have successfully demonstrated that the control flow integrity of processes running on Windows and Linux computers can be reliably monitored [31]. Common integrity breaches caused by CyWs can be identified and countered in real-time.

In the context of AC, the information on the integrity breach (potential CyW) needs to be gathered and collected in a CIP for which reliable trusted communication needs to be established. In particular, the problem of Man-In-The-Middle (MITM) attacks [32] as well as racing conditions between the TPM and the potential malware needs to be solved; we need to establish TNs of trusted machines following the white-listing principle. Peer instances communicating with the system and belonging to the same TN implementing this approach have to be reliably informed about the integrity breach within these TNs. Technically, with the concept of a trusted link-layer protocol establishing secure communication between TPM modules of the peering system, we were able to demonstrate MITM-resistant TNs that are specifically designed to win a racing condition against malware. More specifically, we demonstrated and proved the concept at the example of the Address Resolution Protocol (ARP) – Trusted-ARP (TARP) which we implemented in an extension module for the TPM – an Attack Recognition Module (ARM) [33]. This indicates that we are able to ensure CyW detection by applying the white-listing principle at process level within an entire network segment (and, in theory, with no limits in scalability).

What is needed to make this work? The technical management of such a TN has to be ensured by a local and trustworthy module in every computer that has a secured communication channel to both the communication modules used for exchange of information with peer computers as well as to the internal module detecting integrity breaches. Our suggested approach comprises of technical standard messages to manage peer membership as well as alert message distribution and acknowledgements to ensure reliability. Technically, and from an abstract view, the TN is a link layer network between ARM modules.

From an organisational point of view, FATE-based TNs can be built-up in a similar manner as state-of-the-art networks. A viable option would be to use local networks within organisations, mashed networks connecting these organisations, and, due to their scalability, set up TNs for entire nations. CIPs can then be generated at any granularity level (local network, Internet Service Providers at different TIER-levels, national and international governmental entities or organisations). A CCSP as a combination of CIPs and CAPs allows reliable recognition of CyWs without the need of knowing details of software or data being stored or processed on the affected system. We believe that such a CCSP can be a technical solution for a verification regime to make ACCS work.

5. SUMMARY, CONCLUSIONS AND WAY AHEAD

This paper presented a suggestion for the blue print of an architecture following the white-listing principle to support Arms Control in Cyberspace, based on the requirements derived from concepts and experiences from conventional arms control methods and treaties. A short

description of conventional arms control approaches was followed by a more detailed analysis of four treaties: The Conventional Forces in Europe Treaty (CFE); the Vienna Document (VD); the Open Skies Treaty (OS Treaty); and the Chemical Weapons Convention (CWC). All these agreements point out important aspects when discussing Arms Control in cyberspace, starting by the concept of Treaty Limited Equipment in CFE, followed by the concepts of transparency and confidence-building from VD, followed by the concept of considering the full territorial scope as well as the usage of certified sensors in OS, and finally the discrimination of carrier and payload in CWC.

Based on this analysis, one of the key questions of arms control in cyberspace is whether it is possible to implement a technical approach that allows us to reliably detect any engagement with Cyber Weapons (CyWs). We also showed that trust is a key requirement for any implementation approach. In its technical part, the paper therefore described a trust-based framework based on an extension of the Trusted Computing concept that reliably enables the monitoring and reporting of potential integrity breaches. Based on this mechanism, it is possible to generate situation pictures capable of seeing CyWs in sub-ordinated network structures. We still see research effort necessary within the scope of post-breach activity. Obviously, there are still numerous technical challenges we have to take into account. The points we see as most important to be further elaborated in the future are:

(a) How to react to alert messages/detected integrity violations.
(b) How to defend against Denial-of-Service (DoS) attacks.
(c) How to integrate this concept into operating systems and application processes in a user-friendly manner.

We also have to consider the practical obstacles. For example, in case of a potential treaty violation in CWC, a challenge inspection would be a suitable instrument to prove or disprove the presence of a chemical weapon. What would be the suitable follow-up activity for a detected integrity-breach? Recently we spent some research effort on digital forensics techniques, which we believe to be another very useful enhancement of the Trusted Computing technology and a useful instrument for arms control in cyberspace. As a possible solution, we can collect evidence of the CyW engagement, digitally sign it, and export it to an inspection-site for further investigation [33].

Of course, being familiar with the technical concepts of the Future Advanced Trusted Environment (FATE) we introduced in this paper, this mechanism can also be abused to launch DoS-attacks against FATE-protected systems. We discussed possible solutions for such kinds of scenarios in [34], but further research on this issue is required; we will elaborate on this in more detail in a separate publication. The biggest technical obstacle for successful implementation of our suggestion is the need to adapt operating systems to this new technology. We can implement the required enhancements to operating systems such as Linux since source codes are available, but what about the big proprietary operating systems? Will big vendors such as Microsoft or Apple be willing to support such a new standard?

Our main argument to further promote the idea of our FATE-architecture is the fact that we

can measure integrity without the need of knowing details of the system, either of the software installed or the data being processed. The confidentiality of system contents makes us believe that the necessary political will for an ACCS regime may be achieved. We feel that the concepts can be demonstrated to both the technical thought leaders and political decision makers around the globe and that the core ideas of our proposal will be understood. We also think that the system is not necessarily in conflict with other possible intentions of a state (e.g. active cyber operations), which in daily political business can also be a blocking point – depending on the priority of interests. We think that ACCS can be a next successful step in the history of arms control. To make it a story of success, we will continue developing the FATE framework by designing a software prototype as a proof-of-concept.

REFERENCES

[1] Paletta D., Yadron D. and Valentino-Devires J., Cyberwar Ignites a New Arms Race, in: Wall Street Journal – web portal, October 2015. Available: http://www.wsj.com/articles/cyberwarignitesanewarmsrace1444611128.

[2] Govan G. G., Conventional Arms Control in Europe: Some Thoughts About an Uncertain Future, in: Deep Cuts Issue Brief #5 – Conventional Arms Control in Europe, July 2015. Available: http://deepcuts.org/files/pdf/Deep_Cuts_Issue_Brief5_Conventional_Arms_Control_in_Europe%281%29.pdf.

[3] Van Puyvelde D., Hybrid war – does it even exist?, in: NATO Review magazine, 2015. Available: http://www.nato.int/docu/Review/2015/Also-in-2015/hybrid-modern-future-warfare-russia-ukraine/EN/index.htm.

[4] Denning D. E., Obstacles and Options for Cyber Arms Controls, presented at Arms Control in Cyberspace, Heinrich Böll Foundation, Berlin, Germany, June 2001. Available: http://faculty.nps.edu/dedennin/publications/berlin.pdf.

[5] Arimatsu L., A Treaty for Governing Cyber-Weapons: Potential Benefits and Practical Limitations, in: Proceedings of the 4th International Conference on Cyber Conflct, June 2012. Available: https://ccdcoe.org/cycon/2012/proceedings/d3r1s6_arimatsu.pdf.

[6] Nye J. S. Jr., The World Needs New Norms on Cyberwarfare, in: Washington Post – web portal, October 2015. Available: https://www.washingtonpost.com/opinions/the-world-needs-an-arms-control-treaty-for-cybersecurity/2015/10/01/20c3e970-66dd-11e5-9223-70cb36460919_story.html.

[7] The Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies – web portal, 2016. Available: http://www.wassenaar.org/.

[8] Arquilla J., From Russia Without Love, in: Communications of the ACM, June 2015. Available: http://cacm.acm.org/blogs/blog-cacm/187854-from-russia-without-love/fulltext.

[9] Sanger D. E., U.S. and China Seek Arms Deal for Cyberspace, The New York Times, 19. September 2015. Available: http://www.nytimes.com/2015/09/20/world/asia/us-and-china-seek-arms-deal-for-cyberspace.html.

[10] Manual on International Law Applicable to Air and Missile Warfare, 2009. Available:http://ihlresearch.org/amw/HPCR%20Manual.pdf.

[11] Schneider B., Definition of 'Weapon of Mass Destruction', in: Schneider on Security, 2009. Available: https://www.schneier.com/blog/archives/2009/04/definition_of_w.html.

[12] Boothby W. H., Methods and Means of Cyber Warfare, in: International Law Studies – U.S. Naval War College, Vol 89, 2013. Available: https://www.hsdl.org/?view&did=734393.

[13] Organisation for Security and Co-operation in Europe, Treaty on Conventional Armed Forces in Europe, November 1990. Available: http://www.osce.org/library/14087.

[14] Organisation for Security and Co-operation in Europe, Vienna Document 2011, December 2011. Available: http://www.osce.org/fsc/86597.

[15] Organisation for Security and Co-operation in Europe, Agreement on Adaptation of the Treaty on Conventional Armed Forces in Europe, November 1999. Available: http://www.osce.org/library/14108.

[16] Russia Today, Russia completely ending activities under Conventional Armed Forces in Europe Treaty, March 2015. Available: https://www.rt.com/news/239409-russia-quits-conventional-europe/.

[17] Arms Control Association, Vienna Document 1999, in: Arms Control Association – web portal, August 2010. Available: https://www.armscontrol.org/factsheets/ViennaDoc99.

[18] Organisation for Security and Co-operation in Europe, Forum for Security Co-operation, in: OSCE web portal, January 2016. Available: http://www.osce.org/fsc.
[19] NATO Cooperative Cyber Defence Centre of Excellence, Cyber Definitions, January 2016. Available: https://ccdcoe.org/cyber-definitions.html.
[20] Geers K., Strategic Cyber Security, CCD COE Publications, 2011.
[21] Organisation for Security and Co-operation in Europe, Treaty on Open Skies, March 1992. Available: http://www.osce.org/library/141278.
[22] Gupta, A., The AntiVirus is Dead – says Symantec, The Windows Club Tech News – web portal, May 2014. Available: http://news.thewindowsclub.com/antivirus-dead-says-symantec-68180/.
[23] Organisation for the Prohibition of Chemical Weapons, Chemical Weapons Convention, April 1997. Available: https://www.opcw.org/chemical-weapons-convention/.
[24] Arms Control Association, The Chemical Weapons Convention (CWC) at a Glance, in: Arms Control Association – web portal, October 2015. Available: https://www.armscontrol.org/factsheets/cwcglance.
[25] Organisation for the Prohibition of Chemical Weapons, OPCW Member States, March 2016. Available: https://www.opcw.org/about-opcw/member-states/.
[26] Brangetto P., Caliscan E. and Maybaum, M., Responsive Cyber Defence, study & workshop presentation, presentation at the 6th International Conference on Cyber Conflict, CyCon 2014. Available: https://ccdcoe.org/cycon/2014/.
[27] Ziolkowski K., General Principles of International Law as Applicable in Cyberspace in Zilkowski, K., Peacetime Regime for State Activities in Cyberspace, NATO CCD COE Publications, 2013.
[28] Organisation for Security and Co-operation in Europe, Confidence-building measures to enhance cybersecurity in focus at OSCE meeting in Vienna, November 2014. Available: http://www.osce.org/cio/126475.
[29] United Nations, The Biological Weapons Convention, March 2016. Available: http://www.un.org/disarmament/WMD/Bio/.
[30] Trusted Computing Group, Trusted Computing - web portal, 2016. Available: http://www.trustedcomputinggroup.org/trusted_computing.
[31] Maybaum, M., Trusted Control Flow Integrity, in: Risiken kennen, Herausforderungen annehmen, Lösungen gestalten, SecuMedia Verlag, Gau-Algesheim, Germany, May 2015. (in German)
[32] IBM, Man-in-the-Middle, Trusteer web portal, 2016. Available: http://www.trusteer.com/en/glossary/man-in-the-middle-mitm.
[33] Maybaum, M. and Toelle, J., Trusted Forensics, to be published in: Proceedings to the 15th European Conference on Cyber Warfare and Security, ECCWS 2016, July 2016.
[34] Maybaum, M. and Toelle, J., ARMing the Trusted Platform Module, in: Proceedings to IEEE Military Communications Conference, MILCOM 2015, October 2015.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Malware Counter-Proliferation and the Wassenaar Arrangement

Trey Herr
Belfer Center's Cyber Security Project
Harvard Kennedy School
Cambridge, M.A., USA

Abstract: Can states control malware? It is a radical asymmetry: the power of a modern nation-state arrayed against a few hundred thousand bytes of code and its creators, but an imbalance whose counterintuitive nature impacts the security of citizens, corporations, and governments alike. This paper evaluates export controls found in the Wassenaar Arrangement targeting malware in the context of the research into the malicious software ecosystem. The article highlights that export controls place burdensome restrictions on research and commercial information security efforts. Looking at the market for malicious software, incentivising the discovery of new vulnerabilities to reduce their availability to attackers would be more effective in curtailing malicious activity.

Keywords: *foreign policy, proliferation, cyber crime, malware market, Wassenaar Arrangement, export controls*

1. INTRODUCTION

The malicious software ecosystem is built on a set of interlinked markets where a combination of exploits, payloads, and more developed services like botnets are bought, sold and rented. Unlike more traditional illicit goods markets, the principle content of these products is information. The value of goods in the malware market is thus often derivative of their secrecy, as each can be replicated without marginal cost.

Government's interaction with these markets and the application of state power to restrict the flow of goods into and out of this malware ecosystem has been a public policy issue for many years. However, international efforts to curtail the deployment of malware and its growth in sophistication have yielded limited success. The controversy over changes to the Wassenaar

Arrangement, while an intriguing illustration of state-civil society relations and the nature of malware, obscures a deeper misunderstanding of the threat environment.

This paper leverages previous research into malicious software markets and a review of internal documentation from intermediaries and broker firms in this ecosystem. These different sources are combined to present a picture of the malware market's structure and its interaction with state power. By evaluating this market against the malware-specific language of Wassenaar, this paper highlights three key limitations in the export control approach to counter-proliferation – harm to beneficial research activity, minimal impact on malicious activity, and the challenge of creating multilateral coordination regimes. Increasing the clarity of national policy goals around malicious activity and creating better incentives for vulnerability research represents a more effective path forward.

2. THE WASSENAAR ARRANGEMENT

The Wassenaar Arrangement was signed in 1996 from a desire to revise the Cold War era arms export control process and to integrate former Soviet bloc states into an international forum to restrict the flow, dissemination and proliferation of potentially dual-use technologies to terrorist groups and states of concern (Dursht, 1997). Wassenaar is not an enforceable legal regime or a treaty, but rather a means for participating states to coordinate their respective domestic policies. It is intended to prevent regulatory arbitrage, where businesses relocate their activities to the least restrictive jurisdiction. This arbitrage can undermine export controls and pose a particular challenge for information security (Farrell, 2006).

In December 2013, the British and French governments proposed, with the support of other member states, to change the Arrangement's language and add new rules to cover the tools and technology associated with malicious software (Anderson et al., 2015). One goal for the new controls was to restrict the sale of malware to repressive governments that used technology to monitor journalists and political dissidents. Evidence of this pervasive surveillance industry involved a host of companies in countries like Egypt, Bahrain, and Pakistan (Marquis-Boire et al., 2013; Gamma International, 2013; Hypponen, 2011). The original intent of the new controls was to target firms such as Hacking Team and Gamma Group, which were selling these surveillance tools to these states.

These modifications to Wassenaar created a new restricted product, 'intrusion software', which referred to the tools designed to bypass defences, gain access to computers, and extract data from them. Rather than target these products directly, the Wassenaar language provides for controls on the supporting infrastructure used to generate, deploy, or communicate with this intrusion software (Wassenaar Arrangement, 2015). The controls discussed here are a subset of the broader Wassenaar changes that include separate restrictions on IP network surveillance tools intended for the collection and analysis of large volumes of network traffic. These products are distinct from malware and thus not considered here.

There are additional subtleties around the construction and use of these malware components that are not detailed here. Exploit development, for example, involves a discovery and testing process to advance from knowledge of a vulnerability to an exploit suitable for use on a target system, not just a basic proof of concept. This process takes time, talent, and a not insignificant measure of luck. The persistence of these components, a feature that may differentiate state from non-state group's code, tends to originate in the design of the respective payloads and propagation methods and deserves a more comprehensive treatment than is possible here.

The malware specific changes to Wassenaar were not directly instigated by the United States, but much of the ensuing controversy was provoked by their implementation in American export control law which broadens the original language (Bureau of Industry and Security, Department of Commerce 2015). In May 2015 the US Department of Commerce issued a draft rule which expanded the rules to require source code submission with license applications and proposed to apply a policy of 'presumptive denial' for products which integrated zero-day vulnerabilities (Fidler, 2015). These controls attempted to replicate the sort of restrictions US export law placed on cryptographic tools, targeting their sale rather than use.

3. THE MARKET FOR MALWARE

A. What is malware?

There are a host of different ways to talk about malicious code and categorise its various functions and features; perhaps too many, as competing definitions and naming schemes abound. To focus the discussion, this paper employs the PrEP framework which describes malicious software tools as sharing three fundamental components: a propagation method, exploits, and a payload (Herr, 2014).

The propagation method is a means of transporting code from its origin to the target computer. Anything that can hold or transmit data can propagate malware, be it an email attachment or USB memory stick. Exploits are code designed to take advantage of software flaws and enable attackers to compromise a computer system; they support both the propagation method and payload. The payload contains the malware's key functionality: software delivered to a computer to achieve some specific goal, such as pilfering intellectual property or causing physical damage. The botnet infrastructure employed as a propagation method in spear phishing attacks, for example, is quite different from the payload these attacks distribute, and they are often bought and sold separately. It is important to note that while this framework can be a useful conceptual tool, it is not intended for direct adoption as a legal construct.

B. Why a market?

The market is a well-studied phenomenon in social science. Using the models and language of a marketplace allows us to tie in existing scholarship and to structure an analysis of how malicious actors might respond to different policy interventions. The price a state is willing to pay for a certain vulnerability may set the market for other players. This could then encourage new suppliers to develop exploits for the vulnerability and potentially price out relatively

'friendly' actors such as a responsible software vendor. The National Security Agency (NSA) bidding on a vulnerability for Google's Chrome browser might block it from being sold to Google directly and patched.

The existence of a market-like apparatus for the exchange, purchase, and sale of malware components is not a novel idea, with previous work looking at aspects of pricing (Ablon, Libicki, and Golay, 2014) and the commoditisation process (Grier et al., 2012), as well as larger economic structures (Huang et al., 2015). The buyers and sellers are individuals, firms, and even governments (Fung, 2013). Specialisation has become nearly a de-facto standard with one paper observing that 'the underground economy has evolved into a complex ecosystem with commoditised services' (Huang et al., 2015).

Understanding the functioning of this market is important to evaluate how export controls may be applied to limit the type of goods offered and curb the behaviour of participants. This market is largely a notional one – the interactions of vulnerability discoverers, exploit brokers, and buyers occur in a variety of forms and fora across time and space (Motoyama et al. 2011). Employing the framework of a marketplace is made possible by the competition between different sites and suppliers to attract business and generate revenue. Buyers look across a variety of sources to find commodities and services and sellers will concentrate expertise into particular products to improve economies of scale.

C. Market forces

(i) Supply

There is a tremendous variety in the underground forums where buyers select malware components and related services, but also a clear distinction between the supply and demand relationships for vulnerabilities. Demand is determined by buyers with an interest in purchasing and deploying malware. The source of supply varies by malware component. Vendors' software is the primary source of new vulnerabilities, for example. Propagation methods are numerous; an example is the herds of infected computers known as botnets. Marshalling these machines and renting them is not a trivial process and may depend on the skill and pre-existing resources of the seller. Payloads are largely developed by individuals or small groups and may be purchased, modified, or even stolen. The variation in their purpose means that groups with different skill levels may be unsuccessful at adapting the tools of others or creating their own for complex tasks.

Malware components, and in particular payload, may be reused but adapting payloads may require a high degree of engineering effort. Adapting highly specific payloads, like Stuxnet, is difficult. Code written for a more general purpose is easier. A prominent example of this is Zeus, a popular malware family used to target banks and financial institutions, whose source code was leaked online in 2011 without cryptographic protection (Fisher, 2011). Less than six months later, several new malware families had integrated some or all of the Zeus code and saw sales growth as a result (Damballa, 2011).

Malware sales may be for single components or for several assembled into a service like an exploit kit, which combines propagation methods and one or more exploits (Clarke, 2013). The Angler exploit kit, for example, appears to have integrated the Cryptowall payload with great success (Mimoso, 2015a). Groups may also supply fully operational tools like the GammaGroup's FinSpy surveillance package (Anderson, 2015).

(ii) Demand

The determinants of demand vary, as some groups appear to select targets based on the tools available to them while others find tools to satisfy strong preferences about targets. The implications for tool-focused groups are that demand is shaped by the security posture and configuration (e.g. vulnerabilities present and unpatched) of the potential victims. Targets are selected based on the tools available, making these attacks more opportunistic. A different calculus reigns where the potential target is more important than the tool employed. Target focus implies a willingness to develop or to purchase and modify components to fit the needs of a narrowly defined target set. Being target-focused may require groups to dedicate greater resources to develop appropriate code.

(iii) Reputation and trust

How do you establish trust in a den of thieves? Reputation mechanisms are the means by which two parties in a transaction establish a basis of trust on which to guarantee the desired exchange of goods or services for compensation (Yip, Webber and Shadbolt, 2013). Understanding the potential for fraud in any given transaction, buyers must evaluate the relative quality and character of the good being purchased. There are means to overcome this natural information asymmetry. Independent crowd sourced mechanisms can be used to provide evaluations and customer feedback, like the star based ratings systems found on mainstream e-commerce sites such as Amazon and eBay (Holt, 2012).

In the malicious software market, some forums allow users to post information about the quality of code and services received, to complain about poor customer support, or to call out fraudulent transactions or failure to receive the promised product. There are also systems relying on an interpersonal 'vouching' protocol, not unlike friend of friend chains in other illicit environments, which allows existing trusted networks to add nodes at the edges by brokering introductions between previously disconnected parties (Motoyama et al., 2011). Prolific suppliers can use this as a tool to enhance their legitimacy in new environments or even help suppress the sales of competitors as their reputation precedes them.

D. Supply side actors

(i) Malware components and service vendors

Supply side actors are firms and organisations hawking particular malware components, such as a payload focused on the extraction of user credentials or exploit kits. Suppliers in the high end of the market form a highly fragmented ecosystem of skilled individuals who focus on the development and sale of new exploits for well secured software or high value targets. Companies selling exploits, like ReVuln and Exodus Intelligence, generally operate on

a subscription service model comparable to a data plan for a cellphone (Constantin, 2012). In these programmes, governments and other intermediaries pay a certain amount every year in exchange for a fixed number of exploits. By one estimate, in the entire marketplace there are at least half a dozen such firms capable of selling more than 100 new exploits a year to both governments and non-state actors, with an average list price ranging from $40,000 to $160,000 (Frei, 2013).

Suppliers at the low end of the market sell basic malware components like payloads and propagation methods as well as vulnerabilities. Rarely if ever does this low end market see sales of vulnerabilities unknown to the vendor (zero-days). These suppliers also offer services like click fraud, intended to drive visitor traffic through particular ads, and pay-per-install (PPI), where threat groups can pay according to the vendor's success at infecting different users (Team Cymru, 2011). Interaction tends to take place via Internet relay chat (IRC) or through a shifting collection of forums like Agora, Darkode, and Abraxas.

(ii) Supply side intermediaries

Supply side intermediaries are firms with a legitimate business presence who sell malicious software components to buyers, generally states. These groups may participate directly in the market as buyers in order to expand their wares, but also function as suppliers to at least a portion of the market. These intermediaries resell more developed products intended largely for states, but which may also involve some more sophisticated criminal groups. This mix is interesting as some state organisations may be less capable than non-state actors. The canonical state threat, America's NSA or Israel's Unit 8200, is generally accorded a high degree of technical capability. Others, like intelligence bodies of the Republic of Sudan and Ethiopia, appear to be lacking even the skillset required to operate Hacking Team's products without potentially glaring errors and remedial training (Currier and Marquis-Boire, 2015).

Hacking Team, a prominent Italian company whose internal email system and documentation were leaked onto the web in May 2015, uses a regular stream of new exploits to support its core malware called Galileo RCS or Remote Control System (Anderson, 2015). This malware can be used to spy on a variety of desktop and mobile phone operating systems. Customers can then purchase a licence to increase the number of computers and phones they collect information from at any given time. Companies even offer training, customisation and maintenance features for more developed users.

E. Demand side actors

(i) States

States are highly resourced actors with specific objectives beyond just affecting information systems. States are capable of employing, and have demonstrated interest in, destructive payloads. They also have the human capital and time to develop components internally rather than exclusively through purchase or reuse. The capabilities of the most advanced states are moving ahead of the market and the increasing frequency with which they use malicious software creates a proliferation challenge (Herr and Armbrust, 2015). This proliferation disseminates more sophisticated components to other threat actors where they can be discovered

and potentially re-engineered. Nevertheless, this again remains a challenging task depending on the code obtained (Moshe and Keith, 2015).

(ii) Non-state actors

Non-state actors are the canonical information security threat – criminal groups, gangs and malicious individuals. Varying levels of resources and capability mean that a common decision-making process for these groups is difficult to generalise, but they almost certainly have access to fewer human or material resources than states. Non-state groups' interactions are rarely overt, but their goals are generally oriented around financial gain. This category may include organisations with a political or otherwise ideological agenda, but the sophistication of the tools employed tends to be lower. For example, non-state groups will often use distributed denial of service (DDoS) attacks for their disruptive and propaganda benefits.

(iii) Demand side intermediaries

Demand side intermediaries are groups that contribute to victim security. This may be through providing information assurance resources or managed security services. These intermediaries could also be involved in directly purchasing vulnerabilities like Verisign's iDefense Vulnerability Contributor Program (VCP), which buys vulnerability and exploit information with the intent of disclosing it to vendors (Frei, 2013). This disclosure only takes place after a delay, sometimes substantial, in which subscribers to the iDefense program have exclusive access to the information. These defensive firms participate in the market and affect the stock of goods by disclosing vulnerabilities back to vendors after a delay. There is also a small industry of companies who, for a fee, will conduct mock attacks on organisations' networks to pinpoint weaknesses. These penetration testing firms employ exploits in the same manner as a criminal or state attacker, and often have an interest in the latest research in order to be most effective but can purchase exploits more cheaply since they do not require exclusive access.

(iv) Software vendors

Vendors may also be buyers in the market through vulnerability purchase programmes (VPPs), not participating directly but affecting the incentives of potential suppliers and the available price and quality of goods. These take a variety of forms and many reward more than standalone vulnerabilities; compensation is also given out for novel attack and defence techniques. Firms such as Google and Facebook organise 'bug bounties' (also known as Vulnerability Reward Programmes), designed to encourage researchers to disclose vulnerabilities directly in return for prestige and a cash reward (Popper, 2015). Some companies, like HackerOne and BugCrowd, sit in between organisations and vulnerability researchers, managing these bounty programs for software firms (Ellsmore 2013).

There are also competitions where researchers are given set amounts of time to find vulnerabilities in major commercial software, and to prove their effectiveness with a rudimentary exploit. In 2015 Pwn2Own, a competition held in Canada, paid out prises totalling more than $400,000 for vulnerabilities in the Chrome, Safari, Internet Explorer, and Firefox browsers as well as other software (Goodin, 2015). However, the sponsor of a similar competition in 2016, Hewlett Packard, pulled out after concerns that changes to Wassenaar might impose penalties or unmanageable legal costs on the event (Mimoso, 2015b).

4. ANALYSIS

Wassenaar's impact on the malware market affects few supply side actors and generally only those demand side actors contributing to enhanced software security. The Arrangement's language takes an overbroad, and thus largely ineffective, approach to disrupting the global flow of malicious software. It does little to affect the existing malware markets, it harms security research, and its export controls are weak in the face of the difficulties of effective international coordination.

A. Missing the mark

The Wassenaar's language defines 'intrusion software' as:

> 'Software' specially designed or modified to avoid detection by 'monitoring tools', or to defeat 'protective countermeasures', of a computer or network-capable device, and performing any of the following:
>
> [a] The extraction of data or information, from a computer or network capable device, or the modification of system or user data.
> or
> [b] The modification of the standard execution path of a program or process in order to allow the execution of externally provided instructions (Wassenaar Arrangement, 2015)

The Wassenaar controls do not target this 'intrusion software' directly. Instead the controls focus on supporting components which are any software, systems, equipment, components, or technology used to generate, operate, deliver, or communicate with intrusion software. In effect, Wassenaar targets the means by which intrusion software is built, deployed, or communicated with (Dullien, Iozzo, and Tam, 2015).

One of the major sources of innovation on non-state-authored malicious software comes from state-built code. Duqu, a likely state-built espionage platform discovered in 2011, used an exploit in the Windows operating system to escalate privileges on a target machine and enable payload execution (Bonfante et al., 2013b). Less than a year after the announcement of its discovery, the same exploit was integrated into two major exploit kits and used in attacks against a range of targets by criminal groups (Wolf, 2013). This reuptake of the originally state-authored package drove renewed interest in kernel level exploits to the point where Microsoft was forced to more frequently publish patches to this and related vulnerabilities for more than a year afterwards (Mimoso, 2013). Wassenaar fails to affect state actors responsible for some of the latest malware components, especially exploits.

Wassenaar's controls also appear to miss supply side intermediary sales. VUPEN, a French company which sold customised exploits to clients, announced that it would restrict products and sales because of the changes to Wassenaar but although the firm was subsequently removed from the national business registry, the founder and others have gone on to create a new company

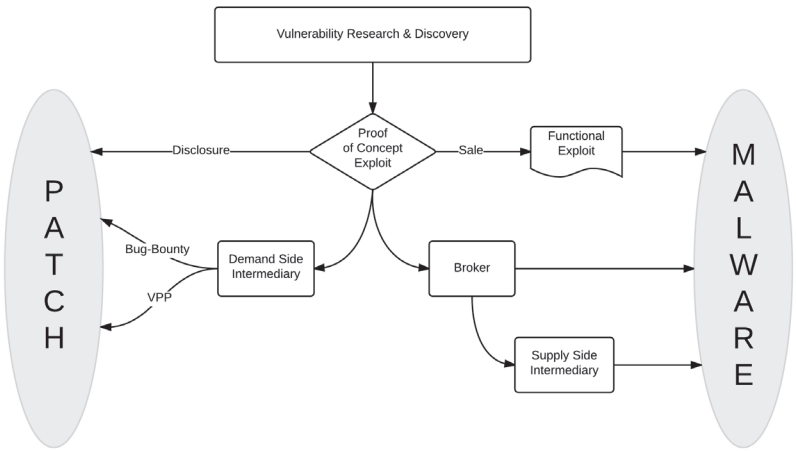
with a similar business model (Granick, 2014). It is not yet clear if any substantive change to the type of malware components being offered or their end use has taken place. In addition, Hacking Team's surveillance malware has a command and control infrastructure, RCS Console, which meets the definition of supporting technologies for intrusion software (Anderson 2015). However, even after Italy implemented new export rules in line with the Arrangement, the firm experienced only a brief interruption in operation (Internet Association, 2015).

B. Collateral damage

Wassenaar's language poses the risk of considerable collateral damage as it might be most effective in targeting organisations contributing to software security and standard software development practices. A vulnerability can be used for ill, by malicious actors, or for good, to patch and improve software security (Shepherd, 2003). There are a variety of actors searching daily for new vulnerabilities, contributing to what has become an arms race of sorts between these researchers and vendors trying to secure their software. Many vulnerabilities are found and quietly fixed through code review by teams housed within giant software vendors like Google, Microsoft, and Adobe. Smaller security firms, academic groups, and independent security researchers play a crucial role as well. These groups often bring new vulnerabilities to light through independent audits, hacking competitions, and bug bounty programs.

Exploits are not intrinsically malicious and thus have limited effectiveness as a signal of the user's intent. They have an array of potential uses in the security industry, and are the principle means by which software vendors are made aware of holes in their products that need to be patched. As with restrictions on cryptography, export controls on malware not only struggle to achieve their goal of restricting the flow of malicious tools around the world, but also create challenges to legitimate users and security research. This is especially true under the current wording of technology for the development of intrusion software (Wassenaar Arrangement, 2015).

FIGURE 1: VULNERABILITY DISCOVERY AND DISCLOSURE



Bypassing protective counter-measures, those so effective as to have become commonplace, is a standard part of security research and experimentation, and exactly the manner in which more secure software is developed. This chilling effect of Wassenaar on research could be substantial as 'nobody can confidently state that he knows how this will be interpreted in practice' (Dullien, 2015). Figure 1 presents a waterfall diagram showing the potential disclosure path of a vulnerability. The Wassenaar rules do as much (if not more) to disrupt the flow of vulnerability information to the patching process as to the malware market.

The Arrangement's language also captures too many potential software tools and security processes to be of any use without creating an untenably complex rule with loopholes and endless exceptions. It isn't that the definition of intrusion software found in the Arrangement will not encompass malware payloads or exploits, but that so much more is swept up at the same time. 'Modification of a standard execution path…', for example, could also include patches from a software vendor to improve user's security, or software plugins like Firefox's Add-ons, which, 'interleave externally provided instructions with the main Firefox code logic' (Bratus et al., 2014). The controls are written in such a blunt manner as to cover quotidian software engineering and security tools, the same sort of broader negative effects on cryptographic tools that began in the 1990s.

C. Challenges to multilateral counter-proliferation

The use of export controls for counter-proliferation of information products has substantial limitations. For both cryptographic tools and malware, one of the key obstacles to successful regulation is that all states party to an agreement must collaborate to prevent regulatory arbitrage. The lack of a standard enforcement mechanism as part of Wassenaar makes this collaboration more unlikely, as different countries may be unable to overcome resistance from domestic constituencies without external inducement (Shehadeh, 1999). Wassenaar also lacks a rule forbidding undercutting, where one state grants export licenses for a product denied by another. There is a provision for notification in these instances, but it does not compel corrective action so a company can choose to export from the most permissive jurisdiction without penalty. While Wassenaar's membership is still broader than that of CoCom, its predecessor, the Arrangement still excludes a majority of countries including the United States, United Kingdom, Germany, France, Estonia, and Russia (Lipson, 1999). This includes information security hubs like Israel, limiting the effective scope of regulation.

Stepping back from the Arrangement, any successor multilateral approach has to consider reporting requirements for specific controls, licensing activity, and violations. Such an international agreement must also include means of maintaining compliance and must cover all states with current or potential information security research programmes. The difficulty in designing an institution along these lines is substantial. Such an agreement would need to include effective reporting and compliance mechanisms and have as members all states likely to be home to research and commercial activity of interest. Even if such an agreement was in place, the Internet has provided more than enough capability for individuals involved in producing, reselling, or brokering malware components to live and work almost anywhere.

5. A WAY AHEAD

Export controls are likely not the best vehicle to target the malicious software ecosystem. Targeting the transmission of these goods and services across national jurisdictions is a near hopeless task, and imposing strong production and research limitations is just as certain to curtail beneficial security efforts. The clearest indication of a tool's potential application is in the design of the payload. Security research or penetration testing services do not need to disrupt the integrity of data on target machines or cause physical damage to demonstrate their efficacy. The canonical testing payload initiates the calculator application on target machines, demonstrating the ability to achieve access and execute successfully. Where payloads may be harder to distinguish, the use of large scale propagation methods, such as distribution through a botnet, could also be indicative of a malicious activity (Herr and Rosenzweig, 2015). The exercise of state power over malicious software will continue to be a substantial policy issue, but there are two particular lessons to take from the current debate.

First, what are the goals of using export controls to target malicious software? The discussion around the 2013 changes to Wassenaar have at times embodied human rights concerns while more recently settling on the national security issue (Maurer and Kehl, 2014). The larger discussion of what counter-proliferation looks like in cyber security, or how best to build institutions to facilitate it, has yet to take place. Understanding what the goals of different states are, limiting state to state proliferation or the use of certain tools against particular groups such as political dissidents is one example of an as yet poorly debated trade-off. More can be done to specify goals and match means to those ends.

Second, existing legal tools may be of a limited use when applied to criminal markets and information products. While policymaking is a process of selecting suitable rather than ideal solutions, the gap between these two has grown wide with the debate around malware. Policy practitioners would benefit from a concerted effort to broaden the discussions specific to Wassenaar towards the larger question of how national policy can positively affect efforts to improve citizens' information security and combat crime. This would be complimented by more systematic engagement with the technical security community and civil society whose insight into the malicious ecosystem can help target policy interventions. A balance of interests is optimal, but some form of state action is inevitable. Continued exchange of common knowledge and language between policymakers and the technical community should not stop with export controls.

An alternative to directly restrict the demand for or use of malware is to target exploits and shrink the supply available for use by states and criminal organisations. In terms of reducing the scale and sophistication of malware, pushing resources and talent towards the goal of finding and patching vulnerabilities would help drive this information to defenders as fast as, or faster than, to those with malicious intent. The community of information security firms and independent researchers responsible for discovering and disclosing a large portion of vulnerabilities every year labour under uncertain legal frameworks that vary between countries. Provisions of the US Digital Millennium Copyright Act (DMCA), for example, can target the analysis of vendor's

products required to find and prove the existence of bugs, enabling lawsuits to prevent this discovery and discouraging security research (Samuelson, 2001; Adams, 2015). In Belgium, similar laws protecting digital rights management (DRM) systems provide restrictions on the form and content of information that can be disclosed about software systems (Biancuzzi, 2008). A temporary security exemption to the DMCA was passed in 2015, but it will not come into effect for a year (Ellis, 2015). The exception protects the work of researchers looking for vulnerabilities in consumer electronic devices, automobiles, and medical equipment from criminal prosecution under the DMCA.

Changes like the DMCA exception could incentivize greater disclosure to software vendors and would help curtail the supply of vulnerabilities to malicious actions. This emphasis on discovering and patching vulnerabilities would also side-step the tricky question of how to directly restrain state behaviour (like purchasing) in these markets. Vulnerabilities are just information so they can exist in multiple places at once. Aggressively identifying and patching bugs in software is not just a way to secure systems, but a means to limit the lifespan of malware in use by states and non-state actors.

6. CONCLUSION

There are a host of problems with the Wassenaar Arrangement's application to malicious software. In the larger analysis of counter-proliferation, export controls targeting malware are likely to prove a poor tool. Additionally, understanding the nature of the underlying technology is important in crafting policy but the operation of a system is not deterministic on the manner in which it is regulated. This shortfall, expecting that laws should map directly onto the nature of a software tool or the process of its development and use, has made security researchers' interaction with the policy process more difficult.

While limited changes to existing law are a positive step, it would be potentially more useful to further revise these protections for research and, importantly, harmonise their application and interpretation across the international community. Vulnerability research is an international enterprise with a bevy of conferences, academic outlets, and competitions taking place across the world. Encouraging security research and growing the pool of individuals looking to discover and disclose vulnerabilities to vendors, would help shrink the pool available to states and criminal groups. Enabling this research through a clear legal framework across states would encourage broader participation and make it easier for vulnerabilities to be disclosed rather than sold on the malicious software market.

The malicious software market is a collection of actors buying, trading, and sometimes stealing from each other. Of the various goods in question, exploits play a key enabling role. However, this level of importance is true of both defensive efforts to engineer secure software and malicious behaviour to compromise it. Exploits are indistinguishable; they do nothing to signal intent. Export controls, as currently constructed in Wassenaar, target these goods as much as any overtly malicious tool. The diversity of demand side actors in this market, and the relatively low

barrier to entry for the supply side make it a difficult environment to restrict with production- or sales-related regulation. Incentivising research could help drain the pool of vulnerabilities for malicious actors.

Wassenaar's changes targeting malware fail to adequately affect malicious actors, place a harmful burden on security researchers, and expose the limits of multilateral approaches to restrict malware proliferation. Improving the clarity of policy goals on combatting crime and enhancing incentives for vulnerability research and disclosure are likely to be more effective than continued application or revision of export controls. The diversity of government, commercial, and civil-society interests in this discussion demands a careful balance, but there remains as yet under-exploited opportunities for crossover engagement and dialogue.

ACKNOWLEDGEMENTS

Thank you to Allan Friedman, Lance Hoffman, Paul Rosenzweig, Katie Moussouris, Eric Armbrust, Tyler Moore, Paulo Shakarian, Ryan Ellis, Andi Wilson, and Jordan McCarthy for their advice and input.

REFERENCES

Ablon, Lillian, Martin C. Libicki, and Andrea A. Golay. 2014. M*arkets for Cybercrime Tools and Stolen Data: Hackers' Bazaar*. Rand Corporation. http://www.rand.org/content/dam/rand/pubs/research_reports/RR600/RR610/RAND_RR610.pdf.

Adams, Stan. 2015. 'Security Research under the DMCA: A Quest for Flexibility and Certainty.' *Center for Democracy & Technology*. July 1. https://cdt.org/blog/security-research-under-the-dmca-a-quest-for-flexibility-and-certainty/.

Anderson, Colin. 2015. 'Considerations on Wassenaar Arrangement Control List Additions for Surveillance Technologies.' Access. https://cda.io/r/ConsiderationsonWassenaarArrangementProposalsforSurveillanceTechnologies.pdf.

Anderson, Colin, CDT, OTI, HRW, and EFF. 2015. 'Comments on the Implementation of 2013 Wassenaar Arrangement.' RIN 0694-AG49. CDT, EFF, HRW, & OTI. https://cdt.org/files/2015/07/JointWassenaarComments-FINAL.pdf.

Biancuzzi, Federico. 2008. 'The Laws of Full Disclosure.' *Security Focus*. February 26. http://www.securityfocus.com/columnists/466.

Bonfante, Guillaume, Jean-Yves Marion, Fabrice Sabatier, and Aurélien Thierry. 2013. 'Analysis and Diversion of Duqu's Driver.' In *Proceedings of the 2013 8th International Conference on Malicious and Unwanted Software*, 109–15. IEEE. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6703692.

Bratus, Sergey, DJ Capelis, Michael Locasto, and Anna Shubina. 2014. 'Why Wassenaar Arrangement's Definitions of Intrusion Software and Controlled Items Put Security Research and Defense At Risk—And How To Fix It.' Public Comment. http://www.cs.dartmouth.edu/~sergey/drafts/wassenaar-public-comment.pdf.

Bureau of Industry and Security, Department of Commerce. 2015. 'Wassenaar Arrangement 2013 Plenary Agreements Implementation: Intrusion and Surveillance Items.' Federal Register. https://www.federalregister.gov/articles/2015/05/20/2015-11642/wassenaar-arrangement-2013-plenary-agreements-implementation-intrusion-and-surveillance-items.

Clarke, Elizabeth. 2013. 'The Underground Hacking Economy is Alive and Well | Security & Compliance Blog | Dell SecureWorks.' December 26. http://www.secureworks.com/resources/blog/the-underground-hacking-economy-is-alive-and-well/.

Constantin, Lucian. 2012. 'Researcher Finds Over 20 Vulnerabilities in SCADA Software.' CIO. November 26. http://www.cio.com/article/2390147/security0/researcher-finds-over-20-vulnerabilities-in-scada-software.html.

Currier, Cora, and Morgan Marquis-Boire. 2015. 'A Detailed Look at Hacking Team's Emails About Its Repressive Clients.' *The Intercept*, July 7. https://theintercept.com/2015/07/07/leaked-documents-confirm-hacking-team-sells-spyware-repressive-countries/.

Damballa. 2011. 'First Zeus, Now SpyEye. Look at the Source Code Now!' *Day Before Zero*. August 11. https://www.damballa.com/first-zeus-now-spyeye-look-the-source-code-now/.

Dullien, Thomas. 2015. 'ADD / XOR / ROL: Why Changes to Wassenaar Make Oppression and Surveillance Easier, Not Harder.' *ADD / XOR / ROL*. May 25. http://addxorrol.blogspot.com.au/2015/05/why-changes-to-wassenaar-make.html.

Dullien, Thomas, Vincenzo Iozzo, and Mara Tam. 2015. 'Surveillance, Software, Security, and Export Controls.' Public Comment. https://tac.bis.doc.gov/index.php/component/docman/doc_view/299-surveillance-software-security-and-export-controls-mara-tam?Itemid=.

Dursht, Kenneth A. 1997. 'From Containment to Cooperation: Collective Action and the Wassenaar Arrangement.' *Cardozo L. Rev*. 19: 1079.

Ellis, Jen. 2015. 'New DMCA Exemption Is a Positive Step for Security Researchers.' *Rapid7 - Information Security*. October 28. https://community.rapid7.com/community/infosec/blog/2015/10/28/new-dmca-exemption-is-a-positive-step-for-security-researchers.

Ellsmore, Nick. 2013. 'Penetration Testing Market Analysis: Where Is All the Revenue?' *Delling Advisory*. April 5. http://www.dellingadvisory.com/blog/2013/4/5/penetration-testing-market-analysis-where-is-all-the-revenue.

Farrell, Henry. 2006. 'Regulation Information Flows: States, Private Actors, and E-Commerce.' *Annual Review of Political Science* 9 (1): 353–74. doi:10.1146/annurev.polisci.9.060804.162744.

Fidler, Mailyn. 2015. 'Proposed U.S. Export Controls: Implications for Zero-Day Vulnerabilities and Exploits.' *Lawfare*. Accessed December 9. https://www.lawfareblog.com/proposed-us-export-controls-implications-zero-day-vulnerabilities-and-exploits.

Fisher, Dennis. 2011. 'Zeus Source Code Leaked.' *Threatpost | The First Stop for Security News*. May 10. https://threatpost.com/zeus-source-code-leaked-051011/75217/.

Frei, Stefan. 2013. 'The Known Unknowns.' NSS Labs. https://library.nsslabs.com/reports/known-unknowns-0.

Fung, Brian. 2013. 'The NSA Hacks Other Countries by Buying Millions of Dollars' Worth of Computer Vulnerabilities.' *The Washington Post*, August 31. https://www.washingtonpost.com/news/the-switch/wp/2013/08/31/the-nsa-hacks-other-countries-by-buying-millions-of-dollars-worth-of-computer-vulnerabilities/.

'Gamma International.' 2013. Reporters Without Borders. http://surveillance.rsf.org/en/gamma-international/.

Goodin, Dan. 2015. 'All Four Major Browsers Take a Stomping at Pwn2Own Hacking Competition.' *Ars Technica*, March 20. http://arstechnica.com/security/2015/03/all-four-major-browsers-take-a-stomping-at-pwn2own-hacking-competition/.

Granick, Jennifer. 2014. 'Changes to Export Control Arrangement Apply to Computer Exploits and More.' *CyberLaw @ Stanford*. January 15. http://cyberlaw.stanford.edu/publications/changes-export-control-arrangement-apply-computer-exploits-and-more.

Grier, Chris, Lucas Ballard, Juan Caballero, Neha Chachra, Christian J. Dietrich, Kirill Levchenko, Panayiotis Mavrommatis, et al. 2012. 'Manufacturing Compromise: The Emergence of Exploit-as-a-Service.' In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, 821–32. ACM. http://dl.acm.org/citation.cfm?id=2382283.

Herr, Trey. 2014. 'PrEP: A Framework for Malware & Cyber Weapons.' *The Journal of Information Warfare* 13 (1): 87–106.

Herr, Trey, and Eric Armbrust. 2015. 'Milware: Identification and Implications of State Authored Malicious Software.' In *NSPW '15 Proceedings of the 2015 New Security Paradigms Workshop*, 29–43. Twente, Netherlands: ACM. doi:10.1145/2841113.2841116.

Herr, Trey, and Paul Rosenzweig. 2015. 'Cyber Weapons and Export Control: Incorporating Dual Use with the PrEP Model | Journal of National Security Law & Policy.' *Journal of National Security Law and Policy* 8 (2). http://dx.doi.org/10.2139/ssrn.2501789.

Holt, T. J. 2012. 'Examining the Forces Shaping Cybercrime Markets Online.' *Social Science Computer Review* 31 (2): 165–77. doi:10.1177/0894439312452998.

Huang, Kurt Thomas Danny Yuxing, David Wang Elie Bursztein Chris GrierD, Thomas J. Holt, Christopher Kruegel, Damon McCoy, Stefan Savage, and Giovanni Vigna. 2015. 'Framing Dependencies Introduced by Underground Commoditization.' In . http://damonmccoy.com/papers/WEIS15.pdf.

Hypponen, Mikko. 2011. 'Egypt, FinFisher Intrusion Tools and Ethics.' *F-Secure Labs*. March 8. https://www.f-secure.com/weblog/archives/00002114.html.

Internet Association. 2015. 'Comments on BIS Implementation of the Wassenaar Arrangement 2013 Plenary Agreements on Intrusion and Surveillance Items.' http://internetassociation.org/wp-content/uploads/2015/07/Internet-Association-Comments-on-BIS-Implementation-of-Wassenaar-7.20.15.pdf.

Lipson, Michael. 1999. 'The Reincarnation of CoCom: Explaining Post-Cold War Export Controls.' *The Nonproliferation Review* 6 (2): 33–51.

Marquis-Boire, Morgan, Bill Marczak, Claudio Guarnieri, and John Scott-Railton. 2013. 'For Their Eyes Only: The Commercialization of Digital Spying.' Citizen Lab. https://citizenlab.org/storage/finfisher/final/fortheireyesonly.pdf.

Maurer, Tim, and Danielle Kehl. 2014. 'Against Hypocrisy: Updating Export Controls for the Digital Age.' *Cyber Dialogue Conference*. May 2. http://www.cyberdialogue.ca/2013/03/against-hypocrisy-updating-export-controls-for-the-digital-age-by-danielle-kehl-and-tim-maurer/.

Mimoso, Michael. 2013. 'Of TrueType Font Vulnerabilities and the Windows Kernel.' *Threatpost*. July 11. https://threatpost.com/of-truetype-font-vulnerabilities-and-the-windows-kernel/101263/.

Mimoso, Michael. 2015a. 'Evasion Techniques Keep Angler EK's Cryptowall Business Thriving.' *Threatpost*. July 2. https://threatpost.com/evasion-techniques-keep-angler-eks-cryptowall-business-thriving/113596/.

Mimoso, Michael. 2015b. 'Citing Wassenaar, HP Pulls out of Mobile Pwn2Own.' *Threatpost*. September 4. https://threatpost.com/citing-wassenaar-hp-pulls-out-of-mobile-pwn2own/114542/.

Moshe, Sariel, and Jacqueline Keith. 2015. 'Recycle, Reuse, Reharm: How Hackers Use Variants of Known Malware to Victimize Companies and What PayPal Is Doing to Eradicate That Capability | PayPal Engineering Blog.' *PayPal Engineering*. November 19. https://www.paypal-engineering.com/2015/11/19/recycle-reuse-reharm-how-hackers-use-variants-of-known-malware-to-victimize-companies-and-what-paypal-is-doing-to-eradicate-that-capability/.

Motoyama, Marti, Damon McCoy, Kirill Levchenko, Stefan Savage, and Geoffrey M. Voelker. 2011. 'An Analysis of Underground Forums.' In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, 71–80. ACM. http://dl.acm.org/citation.cfm?id=2068824.

Popper, Ben. 2015. 'A New Breed of Startups Is Helping Hackers Make Millions — Legally.' *The Verge*. March 4. http://www.theverge.com/2015/3/4/8140919/get-paid-for-hacking-bug-bounty-hackerone-synack.

Samuelson, Pamela. 2001. 'Anticircumvention Rules: Threat to Science.' *Science* 293 (5537): 2028–31. doi:10.1126/science.1063764.

Shehadeh, Karim K. 1999. 'Wassenaar Arrangement and Encryption Exports: An Ineffective Export Control Regime That Compromises United States Economic Interests, The.' *Am. U. Int'l L. Rev*. 15: 271.

Shepherd, S. 2003. 'Vulnerability Disclosure: How Do We Define Responsible Disclosure?' *GIAC SEC Practical Repository, SANS Inst* 9.

Team Cymru. 2011. 'A Criminal Perspective on Exploit Packs.' https://blog.qualys.com/wp-content/uploads/2011/05/team_cymru_exploitkits.pdf.

Wassenaar Arrangement. 2015. 'The Wassenaar Arrangement On Export Controls For Conventional Arms And Dual-Use Goods And Technologies List Of Dual-Use Goods And Technologies And Munitions List.' http://www.wassenaar.org/wp-content/uploads/2015/08/WA-LIST-15-1-2015-List-of-DU-Goods-and-Technologies-and-Munitions-List.pdf.

Wolf, Julia. 2013. 'CVE-2011-3402 - Windows Kernel TrueType Font Engine Vulnerability (MS11-087).' presented at the CanSecWest, March 8. https://cansecwest.com/slides/2013/Analysis%20of%20a%20Windows%20Kernel%20Vuln.pdf.

Yip, Michael, Craig Webber, and Nigel Shadbolt. 2013. 'Trust among Cybercriminals? Carding Forums, Uncertainty and Implications for Policing.' *Policing and Society* 23 (4): 516–39.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Weapons Systems and Cyber Security – A Challenging Union

Robert Koch
Faculty of Computer Science
Universität der Bundeswehr München
Neubiberg, Germany
robert.koch@unibw.de

Mario Golling
Faculty of Computer Science
Universität der Bundeswehr München
Neubiberg, Germany
mario.golling@unibw.de

Abstract: A broad range of weapons systems are in service in forces all over the world. Nowadays, state-of-the-art weapons systems are deployed beside legacy high-value systems that have been used for decades, and will continue to be used for some time. Modern weapons systems can contain hundreds of thousands of chips; each of these chips can be of a sophisticated design, containing billions of transistors, making highly complex systems-of-systems. Elderly weapons systems' service lives are often extended or their performance enhanced due to reduced budget funds or delays in new procurement. Therefore, aged and state-of-the art systems have to function together, not only from a communications prospective, but also from a complete systems integration point of view. Modern Network Centric Warfare scenarios rely upon all of these systems being well integrated and be able to interoperate. This spans an incredibly complex range of sensors, communications systems, and weapons of various ages, opening up countless attack vectors and presenting severe challenges to weapons systems security. The paper analyses the parties involved in today's battlespace, examines the impact of the weapons systems' ages on IT security, and surveys the critical factors for cyber security. Numerous highly dangerous factors are identified and essential necessities and countermeasures are recommended.

Keywords: *weapons systems, COTS, counterfeit chips, network centric warfare, defence electronics, supply chain, cyber war*

1. INTRODUCTION

Nowadays, weapons systems are overwhelmingly complex systems-of-systems, which greatly complicates the analysis of overall system security and increases uncertainty about vulnerability to cyber attack. In addition, while a weapons system as a whole is regularly built on home soil, or at least in close collaboration with partner nations, the integrity of its components is

difficult and costly to assure. For example, the integrated circuits (ICs) used in the computers and communications systems of weapons systems are typically purchased from a variety of sources, often from the lowest bidder. It is exceedingly difficult to levy additional requirements such as monitoring component fabrication or subsystem assembly without incurring significant additional costs. As a consequence, there are often troubling questions about supply chain security. Concerns about the relocation of production from expensive Western countries to lower-priced facilities in Asia arose in the 1990s. First, the loss of intellectual property was feared, but soon the security of highly classified systems equipped with externally made chips was questioned. Because of this, in the early 2000s, the US Department of Defence (DoD) started to look for options to improve the security of sensitive defence systems. Concerns continued to grow after the bombing of a suspected Syrian nuclear installation by Israeli jets in 2007 during Operation Orchard. Because state-of-the-art radar technology was not able to detect the jets, rumours arose that a back door integrated into some chips had been used to compromise the system. While a back door enables potential unauthorised access, another type of hardware-manipulation is the so-called 'kill switch' which can be used to disable a circuit remotely. The 2007 incident greatly boosted worries about possible kill switches within chips in nations' own weapons systems [1].

In order to reduce the risk, the DoD and the National Security Agency (NSA) funded the Trusted Foundry Programme (TFP), for ensuring 'access to microelectronics services and manufacturing for a wide array of devices with feature sizes down to 32 nm on 300 mm wafers' [2]. The program contains 52 trusted suppliers that can establish a trusted supply chain. TFP was completed in 2013 and 'provides national security and programs with access to semiconductor integrated circuits from secure sources' [2]. The programme is able to provide chips for the most sensitive systems, but the complexity of modern weapons systems does not allow the removal of chips from untrusted sources entirely. On the contrary; an investigation in 2011 indicated that 40% of military systems were affected by counterfeit electronics [3]. While efforts within the DoD have improved the situation since 2011, counterfeit parts and the supply chain risks still remains challenging, as a report of the United States Government Accountability Office (GAO) to Congressional Committees highlighted in February 2016 [4].

The trade in counterfeit parts is an increasing threat, opening up different dangers. Counterfeit parts often do not meet the quality requirements of the real products, increasing the risk of malfunction of a weapons system. Counterfeit parts can also increase the risk of back doors and manipulated circuits being present.

While much research has been done to find and improve techniques for the detection of malicious circuits, new and even more dangerous manipulations are possible, and highly sophisticated attacks can be conducted even *below* the transistor level. A worrying example is a recent demonstration of the realisation of a hardware Trojan *below* gate level of an Intel Ivy chip, shown by Becker et al [5]. In contrast to other manipulation techniques such as integrating hardware back doors at gate level which requires about 1300 gates, the authors changed parts of the dopant polarity, and therefore the changes were not detectable on the wiring layers by traditional tests like fine-grain optical inspections or checking against golden chips. Becker was able to reduce the entropy of the integrated random number generator (RNG) from 128 down to

32 bits, enabling easy attacks when the manipulation is known. Testing procedures of the RNG based on NIST guidance are not able to detect manipulations of the generated random numbers. Some cases of back doors implemented in chips are already known, for example the discussion about the Microsemi ProASIC 3 [6] used in military systems and the Boeing 787 Dreamliner [7]. In hindsight, those unwanted circuits are regularly qualified as undocumented debugging functionality. However, for military and highly secure systems, it makes no difference if a hardware back door – which is virtually impossible to detect – was forgotten accidentally or was inserted by purpose.

The remainder of the paper is structured as follows. First, an overview of important characteristics of today's weapons systems is given in section 2, and the impact of Network Centric Warfare on Cyber Security is discussed. In section 3, threats with respect to weapons system and the battlefield are analysed and discussed, while in section 4, possible and necessary countermeasures to reduce the threat to cyber security of weapons systems are presented. Finally, section 5 concludes the paper, highlighting the key takeaways.

2. WEAPONS SYSTEMS

Today's battlespace is filled with an extensive variety of weapons systems of all ages.

A. Elderly weapons systems

The development, commissioning and operation of a weapons system is very expensive. For this reason, such high-value systems are built to be in service for more than 30 to 40 years. While this is a long period of time, it is often extended even further for financial or procurement reasons: economic crises and budget cuts have affected all nations at one time or another, resulting in the cancellation of numerous procurement projects. In addition, the buying process of weapons systems can be very time-consuming. Because of late changes to specifications or issues during the development process, delays of many years are not uncommon in this sector. For example, it was 17 years from foundation of the 'Eurofighter Jagdflugzeug GmbH' in 1986 to the delivery of the first production aircraft in 2003, and the unit costs rose from the originally planned €33.32 million to €138.5 million by 2012 [8]. Such delays also can greatly contribute to the extension of the service life of a weapons system. Due to such developments, the *average* age of the weapons systems of a military force can be several decades, even in modern western forces. For example, the average age of US Airforce aircraft is 27 years, and some fleets like the B-52 bomber, which entered service in 1955, are much older [9]. Therefore, the expected life expectancy of many elements of the United States Air Force (USAF) will be reached if the equipment is not enhanced by modifications [9].

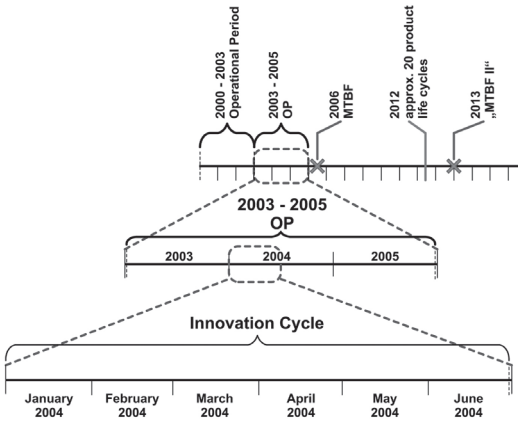
Over the years of operation, the supply of spare parts can also be challenging. Companies might go out of business, change production lines, or produce new and incompatible products. Because of that, and also to maintain availability, mid-life upgrade programmes are conducted one or more times during the life span of a weapons system. More and more commercial-off-the-shelf (COTS) products have to be used to keep systems running and to replace old components which are no longer available [10],[11].

B. State-of-the-art weapons systems

Modern weapons systems are highly complex. They often contain hundreds of thousands of chips, extensive networks, and interconnected sensors and systems. While the military was a driver of technology during the Cold War due to the vast defence budgets for research and development, the peace dividend and several economic and financial crises have necessitated broad budget cuts and reduced overall defence spending. By contrast, the impressive development of Information and Communication Technology (ICT), the Internet and consumer electronics boosted the evolution of the industry to a multi-billion market. Increasingly shorter product and innovation cycles make the commercial market today's driver of technology. To reduce costs as well as to optimise system performance, COTS products are heavily used in modern weapons systems. In turn, the extensive integration of COTS in high-value systems has resulted in new challenges in maintaining the weapons systems over their life cycles. Any attempt to update the ICT components of a weapons system after it is deployed often requires a costly recertification to obtain the authority to operate it. Given the pace of technological advancement, with new ICT being released every few years and weapons systems being designed to operate for a much longer time, it is often cost prohibitive to mandate the update of ICT subsystems in a timely manner.

Looking at the life span of IT components and their mean time between failure (MTBF), and bearing in mind the challenging operational environments such as the broad spectrum of temperatures, material stress caused by high acceleration, or sea disturbance affecting ships, an exchange of these components has to be done at least every ten years. Therefore, a weapons system which is in service for 30 to 60 years requires numerous programmes to refresh it (see Figure 1).

FIGURE 1: HIGH-VALUE SYSTEMS OF THE MILITARY HAVE TO BE IN SERVICE FOR UP TO 40 YEARS OR EVEN LONGER. THEREFORE, ELECTRONIC COMPONENTS LIKE COTS PRODUCTS WITH THEIR TYPICAL MTBF VALUES HAVE TO BE EXCHANGED MULTIPLE TIMES DURING THE LIFE TIME OF THE WEAPONS SYSTEM



MTBF Estimates for Intel Server System SR1630GP

Subassembly (Server in 35oC ambient air)	Server Model SR1630GP MTBF (hours)	FIT (flrs/10^9 hrs)
Intel Server Board S3420GP	208,000	4,808
Power Supply (350W)	400,000	2,500
Cooling fans (Non-redundant)	100,000	10,000
PCI RiserCard	4,100,000	244
Front Panel board	16,000,000	63
Totals without motherboard=	78000	12,806
Totals with motherboard=	56700	17,614

This can cause compatibility issues, and the risk of bringing in manipulated components, counterfeit parts or parts of insufficient quality increases significantly (for example, see [12]).

C. Network-centric warfare scenario

After the end of the cold war, budget cuts forced a more efficient use of funding. While military budgets were reduced, the development, fabrication, operation, and maintenance costs of weapons systems increased steadily. Under these constraints, the best possible use of the limited number of available weapons systems had to be realised. Therefore, the interconnection of all available systems and the appropriate provisioning of any required information at all levels was the answer to maintaining force superiority with a progressively limited, but technically more capable, number of units. This is called Network Centric Warfare (NCW).

The US Navy was one of the first to look at a 21st century battlefield and think of how to use ICT to increase the efficiency of forces [13]. The main consequence of their considerations was the increased integration of individual, previously autonomously acting systems. This technical integration has finally led to the concept of NCW. This is a theory that proposes the application of information age concepts to speed communications and increase situational awareness through networking, and improve both the efficiency and effectiveness of military operations [14]. NCW creates information superiority by means of a network of reconnaissance, command and control, and weapons systems, and thus ensures military superiority across the entire range of military operations (full spectrum dominance). The vision for NCW is to provide seamless access to timely information at every echelon in the military hierarchy. This enables all elements to share information that can be combined into a coherent and accurate picture of the battlefield. This concept of strong and flexible networked military forces allows combat units to be smaller, to operate more independently and effectively, to prevent or reduce fratricide, and to speed up the pace of warfare in comparison to non-networked forces [14]. NCW will also produce an improved understanding of higher command's intent, improved understanding of the operational situation at all levels of command, and an increased ability to tap into the collective knowledge of all forces to finally reduce the 'fog and friction' [14]. While the concept of NCW enables the optimal use of resources, the vulnerability of the overall system rises dramatically: attacking the weakest link of the NCW chain can have catastrophic consequences for its owner, in the worst case rendering a whole military component incapable of action.

3. THREAT ANALYSIS

Having a look at the wide range of weapons systems in today's battlespace and the numerous attack vectors which are connected with them, questions about the most dangerous vulnerabilities arise. In modern warfare, all systems are highly interconnected and gravitate towards NCW scenarios, and a breach of the weakest link can have a severe effect for a whole operation.

A. Old versus new

Because of the mix of elderly weapons systems and those that are state-of-the-art, one might come to the conclusion that older weapons systems are per se more unsecure than newer systems. Although that is often true in that older systems run on older software and may have

challenges with respect to software updates and patches, modern systems have their own problems with outdated software because of the long design and procurement times. Because of mid-life upgrade programmes, old weapons systems are modernised, sometimes with replacement of nearly all their ICT components. Therefore, elderly as well as state-of-the-art systems can contain old as well as new IT components, and have to be treated equally with respect to cyber threats.

B. Defence technological and industrial base (DTIB) capabilities

The capability of the Defence Technological and Industrial Base (DTIB) is another important aspect. The DTIB is the combination of people, institutions, technological expertise, and production capacity used to develop, manufacture, and maintain the weapons and supporting equipment needed to achieve national security objectives [15]. The European Defence Agency defined a strategy for a European DTIB, aiming at strengthening available European capabilities, motivating higher investment, and promoting the broader use of the public procurement regulations of the EU [16]. While Europe has many capable defence companies, components such as electronic semiconductors are often not produced in Europe, but in the Asia-Pacific region [17]. At present, the limited prospects to maintain core IT components of weapons systems are also not addressed explicitly within the DTIB strategy of the EU; in fact, this strongly limits the effectiveness of the European DTIB, and also those of other Western DTIBs that suffer from similar restrictions.

C. Supply chain

While producers in North America slowly stabilise their market share after losing most of it in the nineties and the first decade of this century, Europe's electronic equipment production is still declining, while also China is being challenged by upcoming producers in other Asia-Pacific regions like India and Malaysia [17]. Therefore, the supply chain of IT technology currently presents a severe danger to the security of weapons systems. Because of the complex and globally distributed chip design ecosystem, a multitude of companies and countless people are involved in the building process, from specification to shipping [18]. Driven by the optimisation of business processes, cost-reduction in manufacturing, outsourcing, and globalisation, building a chip nowadays involves a huge number of parties at every step: specification, design, manufacture, and testing. The various steps are distributed between numerous companies. Nowadays, even parts of a chip can be reused or purchased from other companies and the huge number of people involved during chip creation enables a growing threat of design corruption [18].

All steps of the building process can be manipulated to a greater or lesser extent. Some examples of this include: manipulation of specifications; [19] influencing the design process by introducing back doors; forgetting to remove debugging functionality – see the discussion about the hardware backdoor in the Actel/Microsemi ProASIC3 chips; [6] and executing very small changes during chip manufacture (for example, adding a back door requires about 1300 gates, [20] while in contrast, the recent SPARC M7 contains 10 billion transistors – therefore as few as 0.000013% gates have to be added; these are virtually undetectable for today's typical test suites which are able to identify accidental design flaws effectively based on calculus of

probabilities, but which struggle to find intentionally hidden alterations made by a skilled designer [18]).

In addition to these possibilities, an increasing market exists for the sale of counterfeit parts. In 2012, worldwide trade in counterfeit semiconductors reached $169 billion [21]. As this is a lucrative business, a strong further increase can be assumed. Counterfeits endanger weapons systems in two ways: the poor quality typically cannot meet the original specification; and the risk of there being manipulated circuits increases dramatically. In 2012, it was reported that 'a record number of tech products used by the US military and dozens of other federal agencies were fake. That opens up a myriad of national security risks, from dud missiles to defective airplane parts, to cyberespionage' [12]. For detailed examples of supply chain vulnerabilities and resulting risks to DTIB, see [22].

D. Compatibility and maintenance supportability

The long life of weapons systems can also be challenging when components which have to be replaced are no longer available on the market. Often, new products are not compatible with older ones; but even if a newer product is compatible, various problems may occur in practice (for example, while SCSI components should be backwards compatible, it should be possible to use an Ultra-160 SCSI disc on the bus of a SCSI-1 host adapter). Even though this is possible in theory, device compatibility is often reduced in practice, such as when there are different signalling implementations even within the same standard. In the worst case, obsolete components have to be installed in a weapons system to keep it running, increasing costs as well as the danger of counterfeit electronics.

Also, mass market electronics are typically not optimised for radiant emittance further than the basic requirements of electromagnetic compatibility dictate (for example, the EU directive 1999/5/EC on radio equipment and telecommunications terminal equipment [23] or directive 2004/108/EC on the approximation of the laws of the Member States relating to electromagnetic compatibility [24]). With respect to security-related systems, such directives are often not sufficient: for example, electromagnetic compatibility is defined in Article 2 of 2004/108/EC as: 'the ability of equipment to function satisfactorily in its electromagnetic environment without introducing intolerable electromagnetic disturbances to other equipment in that environment'. No threshold values are given by the directives, causing possible trouble when systems have to be replaced but new products cannot fulfil function parameters like the old one.

Today's weapons systems are threatened by fake microelectronics and possible manipulations of the hardware, limited supply of spare parts, and counterfeit parts. While real examples are naturally highly classified, the real-world danger can be recognised by examples like manipulated processors [6], the recent discussion about back doors in widely-used network equipment [25], or statements. For example, IEEE Spectrum wrote in 2008, that:

> 'according to a US defence contractor who spoke on condition of anonymity, a "European chip maker" recently built into its microprocessors a kill switch that could be accessed remotely [...] If in the future the equipment fell into hostile hands, "the French wanted a way to disable that circuit," he said' [1].

4. COUNTERMEASURES

The most important precondition for defining adequate countermeasures is to accept the current situation and the undesirable, but in the medium term unchangeable, reality. It is neither possible to exchange the entire hardware layer to a trusted one, nor to build an all-embracing DTIB. One must act on the assumption that the already deployed hardware-layer is not trustworthy. Therefore, pre-planned reactions for a worst-case scenario are elementary.

A. Emergency planning

The defence capabilities and weapons systems of a nation state are of interest to other nations and non-state actors. The components of today's complex weapons systems and their innermost component parts, especially the chips controlling sensors, communication systems, data exchange and weaponry systems, are often COTS products delivered by a lengthy supply chain involving hundreds of companies and thousands of people. Reactions and countermeasures must be developed and installed on all units and elements as well as at all management levels. This includes the creation of emergency and disaster plans. In order to keep the associated complexity manageable, vital components have to be identified and addressed.

B. Risk management

An organisation-wide risk management must be established, encompassing all units and management levels, but which also has to be integrated in the procurement and planning processes. For this purpose, the standards published by the International Organisation for Standardisation (ISO), can be referred to: ISO 31000:2009 (Risk management – Principles and guidelines) and IEC 31010:2009 (Risk management – Risk assessment techniques). COBIT 5 for risk can be used to implement a holistic and organisation-wide risk management regime. This includes guidance on how to manage the risk to levels, how to implement extensive measures, and how to enable stakeholders to consider the cost of mitigation and required resources as well as other aspects such as setting up an appropriate risk culture [26].

C. Supply chain

Today, specific threats exist not only in the production process of chips, but in particular during the design phase. Therefore, all steps from specification to shipping have to be taken into consideration to establish a more trustworthy supply chain. However, the economic reality, with its globalised business processes, must be accepted and be reflected in an appropriate strategy for dealing with the supply chain. For example, by boosting and funding research for new processes and technologies for securing design tools and development.

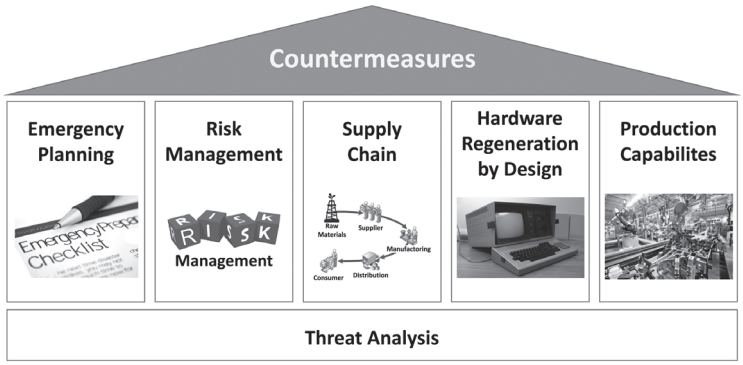
D. Hardware regeneration by design

Looking at individual weapons systems, procurement processes have to be adapted to reflect the requirements of operating short-lived COTS hardware in long-lasting high-value weapons systems. Therefore, not only an exchange of semiconductor components on a regular basis is necessary, but also the provision of concepts of how to deal with compatibility issues, methods for the detection of counterfeit or manipulated chips, and migration strategies for the system core elements in case of severe problems of incompatibility with new hardware.

E. Production capabilities

A strengthening of the European DTIB is necessary to provide all essential components of weapons systems, including the production of semiconductor electronics for the most sensitive systems. This includes a further strengthening of companies already producing crypto- and specialised chips for high-security systems, as well as updating the strategy for the European DTIB, and the creation of an own production capacity, following the example of the TFP.
Figure 2 summarises the identified fields of action which are required to improve the security of weapons systems, and attenuate worst-case scenarios.

FIGURE 2: FIELDS OF ACTION FOR AN IMPROVEMENT OF THE SECURITY OF WEAPONS SYSTEMS AND ATTENUATING WORST-CASE SCENARIOS



5. CASE STUDY: SUPPLY OF SEMICONDUCTORS IN THE US MILITARY

In order to demonstrate the increasing challenges with respect to the supply of semiconductors in US military electronics, this section presents a case study, highlighting the effects achievable by applying the proposed countermeasures. While the US is still a global leader in research and development (R&D) in the semiconductor industry, ever growing proportions of the fabrication take place in the Asia-Pacific region, and this is likely to grow over the coming years. A situation that Brigadier General (ret.) John Adams summarises as follows:

> 'The Chinese telecommunications industry has grown rapidly, with Chinese-manufactured telecommunications equipment spreading swiftly due to below-market prices supported by funding from the Chinese military. The widespread use of military-funded Chinese equipment in conjunction with the shrinking market share of trusted US telecommunications firms increases the likelihood that kill switches or back doors will be inserted into key communications infrastructure, jeopardizing the integrity of sensitive defence-related communications' [22].

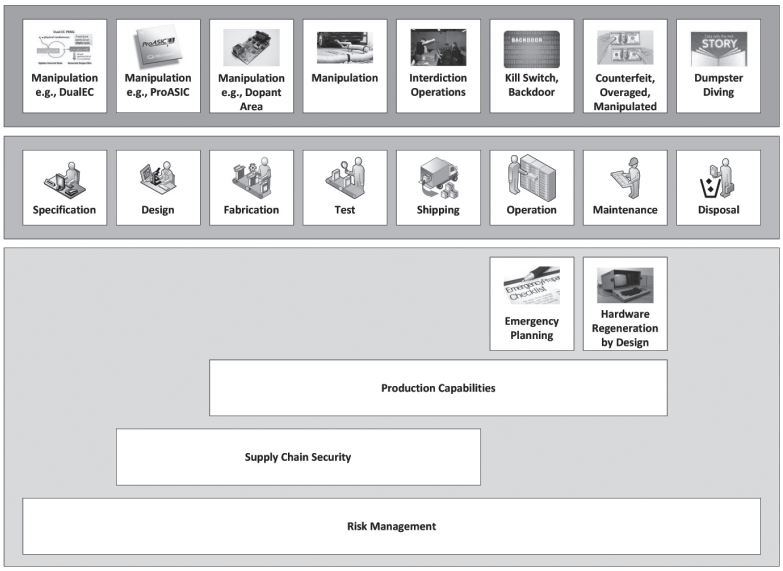
Chinese companies like Huawei or ZTE have a large, constantly increasing market share, providing important network equipment to various customers including military systems and communications. Manipulated circuits containing kill switches or back doors could easily be introduced into systems and networks, but locally designed circuits made by trusted companies using offshore factories can also be infiltrated in that way. For example, telecommunications equipment like the AN/VRC-92A vehicular radio set or the AN/PSC-2 radio may be affected by untrustworthy components (e.g., see [22]). Having a look at the proposed countermeasures, the following measures will be effective in case of manipulated circuitry which may already be used in operational equipment:

(i) Extensive and realistic emergency planning must be able to provide all necessary guidelines to sustain an outage of the respective equipment. For example, if a communication system of manufacturer A is failing (because of the activation of a kill switch), another available device built by manufacturer B and preferably based on different architecture must be able to take over the services. The emergency plan must contain all information to enable the operator to execute all necessary steps (in manifold scenarios) as fast as possible and must be exercised regularly.
(ii) Appropriate risk management must provide guidelines, including the consideration of circumstances that may influence the operation or security. For example, having a manipulated processor within a system where the manipulation cannot be exploited (e.g., an isolated system without any connectivity) does not impose any limitations, while a system connected to the Internet may be highly vulnerable.
(iii) The building of fabrication capabilities can also be used to replace untrusted components of endangered systems that are already in use.

Applying all countermeasures, the risk of introducing manipulated circuitry can be reduced by providing fabrication capabilities for the most essential and restricted systems, and increasing the supply chain security to reduce possible manipulations (e.g., see [27]). Especially for new procurement projects, the necessity of regular hardware regenerations must be incorporated by design. For example, the specification and selection of components must be done in a way that common, long-lasting and open standards (like for example IPv4/6) are used, enabling a replacement of components with minimal adjustment, even when the original supplier is out of business or the product line was phased out. Because of the increased risk of counterfeit parts, novel compatible parts also should be used if original parts are only available from untrustworthy sources.

Figure 3 gives an overview of the obtained effects with respect to the supply chain, when all countermeasures are applied.

FIGURE 3: HOLISTIC RISK MANAGEMENT



6. CONCLUSION

The modern battlefield is a complicated environment where numerous highly complex weapons systems which are a broad mixture from several generations, and variations of electronic equipment of different ages interact. This generates special demands on cyber security, but these sensitive systems are increasingly vulnerable to cyber attacks, as examples like Operation Orchard have shown. Based on the relocation of production capabilities to lower-priced countries and the globalisation of chip production, numerous threats of attacks or manipulations are endangering modern weapons systems. Because of the strongly diverging life cycle times of COTS products and military high-value systems, and the prevalence of counterfeit electronics, the required exchange of hardware components on a regular basis opens up significant challenges. While elderly and modern weapons systems are facing these same challenges and threats, the total risk increases dramatically within a NCW scenario.

Therefore, organisation-wide emergency-management and risk-management is vital. Each unit as well as all management levels must be able to react promptly in case of attack executed at the hardware layer. A strengthening of the European DTIB capabilities especially in the field of semiconductors for sensitive systems is necessary, and new concepts and techniques for improving the security of the supply chain for electronic products from the world market have to be developed. While this will have huge costs, the TFP of the United States shows that building up secure microelectronics manufacturing capacities is viable. Having a look at the

procurement processes, the integration of hardware-regeneration concepts on a regular base is essential.

ACKNOWLEDGMENTS

This work was partly funded by FLAMINGO, a Network of Excellence project (ICT-318488) supported by the European Commission under its Seventh Framework Programme.

REFERENCES

[1] S. Adee, 'The hunt for the kill switch,' *Spectrum, IEEE*, vol. 45, no. 5, pp. 34–39, 2008.
[2] C. Ortiz. Dod trusted foundry program. [Online]. Available: http://www.ndia.org/Divisions/Divisions/SystemsEngineering/Documents/.
[3] J. Mick. US GOA [sic.]: 40 Percent of Defense Supply Chain Damaged by Chinese Parts. [Online]. Available: http://www.dailytech.com/US+GOA+40+Percent+of+Defense+Supply+Chain+Damaged+by+Chinese+Parts/article21937.htm.
[4] M. A. Mak, 'Counterfeit parts - DOD needs to improve reporting and oversight to reduce supply chain risk,' United States Government Accountability Office, Tech. Rep., 2016, gAO-16-236, Report to Congressional Committees. [Online]. Available: http://www.gao.gov/assets/680/675227.pdf.
[5] G. T. Becker, F. Regazzoni, C. Paar, and W. P. Burleson, 'Stealthy dopant-level hardware Trojans,' in *Cryptographic Hardware and Embedded Systems-CHES 2013*. Springer, 2013, pp. 197–214.
[6] S. Skorobogatov and C. Woods, Breakthrough silicon scanning discovers back door in military chip. Springer, 2012.
[7] C. Arthur. Cyber-attack concerns raised over Boeing 787 chip's 'back door'. [Online]. Available: http://www.theguardian.com/technology/2012/may/29/cyber-attack-concerns-boeing-chip.
[8] M. Freund, D. Klager, J. Mallien, and D. Neuerer. Totalschaden mit Ansage. [Online]. Available: http://www.handelsblatt.com/politik/deutschland/ruestungsflops-der-bundeswehr-die-entwicklung-des-eurofighter/8232292-3.html.
[9] D. L. Wood, '2016 index of US. military strength,' The Heritage Foundation, Tech. Rep., 2015. [Online]. Available: http://index.heritage.org/military/2016/resources/download/.
[10] S. Kosiak, Buying Tomorrow's Military: Options for Modernising US Defense Capital Stock. Center for Strategic and Budgetary Assessments, 2001.
[11] L. O. Association. Aging weapons systems. [Online]. Available: http://atloa.org/wp-content/uploads/M1_slides.pdf.
[12] D. Goldman. Fake tech gear has infiltrated the USgovernment. [Online]. Available: http://security.blogs.cnn.com/2012/11/08/fake-tech-gear-has-infiltrated-the-u-s-government/.
[13] D. S. Alberts, 'Information Age Transformation: Getting to a 21st Century Military (revised),' DTIC Document, Tech. Rep., 2002.
[14] C. Wilson, 'Network centric operations: background and oversight issues for congress.' DTIC Document, 2007.
[15] W. Slocombe, 'Adjusting to a new security environment: The defense technology and industrial base challenge - background paper,' US Congress, Office of Technology Assessment, Tech. Rep., 1991.
[16] EDA Steering Board, 'A strategy for the European defence technological and industrial base,' European Defence Agency, Tech. Rep., 2007.
[17] E. Publications, 'World electronic industries 2012-2017,' Electronics.ca Publications, Tech. Rep., 2014. [Online]. Available: https://www.electronics.ca/store/world-electronic-industries.html.
[18] J. Villasenor, 'Compromised by design? securing the defense electronics supply chain,' *Brookings Institution Report*, Nov, 2013.
[19] D. J. Bernstein, T. Chou, C. Chuengsatiansup, A. Hülsing, T. Lange, R. Niederhagen, and C. van Vredendaal, 'How to manipulate curve standards: a white paper for the black hat,' Cryptology ePrint Archive, Report 2014/571, Tech. Rep., 2014.
[20] S. T. King, J. Tucek, A. Cozzie, C. Grier, W. Jiang, and Y. Zhou, 'Designing and implementing malicious hardware.' *LEET*, vol. 8, pp. 1–8, 2008.

[21] L. Dignan. Counterfeit chips: A $169 billion tech supply chain headache. [Online]. Available: http://www.zdnet.com/article/counterfeit-chips-a-169-billion-tech-supply-chain-headache/.
[22] J. Adams and P. Kurzer, Remaking American security: Supply chain vulnerabilities & national security risks across the US defense industrial base. Alliance for American Manufacturing, 2013.
[23] 'Directive 1999/5/ec of the european parliament and of the council of 9 march 1999 on radio equipment and telecommunications terminal equipment and the mutual recognition of their conformity,' 1999. [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31999L0005.
[24] 'Directive 2004/108/ec of the European parliament and of the council of 15 December 2004 on the approximation of the laws of the member states relating to electromagnetic compatibility and repealing directive 89/336/eec,' 2004. [Online]. Available:http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:390:0024:0037:en:PDF.
[25] H. Böck. Juniper wegen Hintertüren in Erklärungsnot. [Online]. Available: http://www.golem.de/news/zufallszahlengenerator-juniper-wegen-hintertueren-in-erklaerungsnot-1601-118457.html
[26] Isaca, *COBIT 5 for Risk*. Isaca, 2013, ASIN: B01A1MXZ30.
[27] Ghadge, Abhijeet, Samir Dani, and Roy Kalawsky.'Supply chain risk management: present and future scope.' The International Journal of Logistics Management 23.3 (2012): pp. 313-339.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

UAV Exploitation: A New Domain for Cyber Power

Kim Hartmann
Otto von Guericke University
Magdeburg, Germany
kim.hartmann@ovgu.de

Keir Giles
Conflict Studies Research Centre
Oxford, UK
keir.giles@conflictstudies.org.uk

Abstract: The risks of military unmanned aerial vehicles (UAVs) being subjected to electronic attack are well recognised, especially following high-profile incidents such as the interception of unencrypted video feeds from UAVs in Iraq and Israel, or the diversion and downing of a UAV in Iran. Protection of military UAV assets rightly focuses on defence against sophisticated cyber penetration or electronic attack, including data link intercepts and navigational spoofing. Offensive activity to counter adversary drone operations presumes a requirement for high-end electronic attack systems. However, combat operations in eastern Ukraine in 2014-16 have introduced an entirely new dimension to UAV and counter-UAV operations. In addition to drones with military-grade standards of electronic defence and encryption, a large number of civilian or amateur UAVs are in operation in the conflict. This presents both opportunities and challenges to future operations combating hybrid threats. Actual operations in eastern Ukraine, in combination with studies of potential criminal or terrorist use of UAV technologies, provide indicators for a range of aspects of UAV use in future conflict. However, apart from the direct link to military usage, UAVs are rapidly approaching ubiquity with a wide range of applications reaching from entertainment purposes to border patrol, surveillance, and research, which imposes an indirect security and safety threat. Issues associated with the unguarded use of drones by the general public range from potentially highly dangerous situations such as failing to avoid controlled airspace, to privacy violations. Specific questions include attribution of UAV activities to the individuals actually directing the drone; technical countermeasures against hacking, interception or electronic attack; and options for controlling and directing adversary UAVs. Lack of attribution and security measures protecting civilian UAVs against electronic attack, hacking or hijacking, with the consequent likelihood of unauthorised use or interception, greatly increases the complication of each of these concerns.

Keywords: *drone, UAV, military, communications*

1. INTRODUCTION

As cyberspace has emerged from being a purely computer based virtual reality to bringing about real life impacts, cyber power has become a vital element of hostile action between states, now including military operations. Cyber power is thus no longer a virtual competence. This paper will discuss a field of activity where cyber power has a direct and immediate effect on the conduct of real-world operations, both civilian and military: the use and exploitation of unmanned aerial vehicles (UAVs).

There has been substantial discussion on the issues associated with UAV use in military operations, especially on the ethical aspects of drone strikes [1]. But the specific issue of UAV security has gained broader public attention due to the use of UAVs in non-military activities.

UAVs are rapidly approaching ubiquity, with a growing range of applications. The benefits of utilising UAVs for inexpensive aerial surveillance and survey have been widely accepted. However, with the broader introduction of UAVs to the civilian market for law enforcement, research and entertainment purposes, a new set of security and safety threats have been unwittingly invited. Specific questions currently unresolved include attribution of UAV activities to the individuals actually directing the drone; technical countermeasures against hacking, interception or electronic attack; countermeasures against UAVs which have already been compromised; and options for controlling and directing adversary or hostile UAVs.

Issues associated with the unguarded use of UAVs by the general public range from potentially highly dangerous situations such as failing to avoid controlled airspace, to privacy violations. The lack of security protecting civilian UAVs against electronic attack, hacking, or hijacking, with the consequent likelihood of unauthorised use or interception, greatly increases the complication of each of these concerns. The increased likelihood of hijacking or interception fosters the risk of abusive and dangerous use by cyber attackers, and complicates the attribution issue. The implications are directly relevant to the full range of UAV operations, from use in state-on-state conflict through civilian and law enforcement applications, to simple entertainment use.

This paper explores the use and exploitation of UAVs as a means of implementing cyber power for real-world effects. It discusses why UAVs are targets for cyber actors; how these actors may use UAVs in combat and civilian scenarios; and examples of how UAVs have been exploited in the past through cyber means. It highlights that cyber power as exercised against UAVs demonstrates how cyber competence may be linked to the success or failure of real life combat missions. The paper is written as an aide to policy-makers; an essential technical overview of the range of possible cyber attacks on UAVs is therefore included, but detailed analysis of attacks is not. Instead, the paper aims to provide an introduction to the range of policy implications of the current state of development of UAV security, based on implementation (or lack of it) to date.

2. UAV PAST INCIDENTS

Electronic attacks on UAVs are not new; but while earlier attacks were relatively rare, fairly sophisticated, and directed against military devices due to their tactical, strategic and monetary value, more recently a series of incidents against and/or involving civilian drones have been reported. The latter reflects the recently gained popularity of UAVs for recreational uses, and the resulting potential for abuse. It will be observed that many of these incidents were only possible due to massive flaws in the implementation of security measures, if these measures were implemented at all.

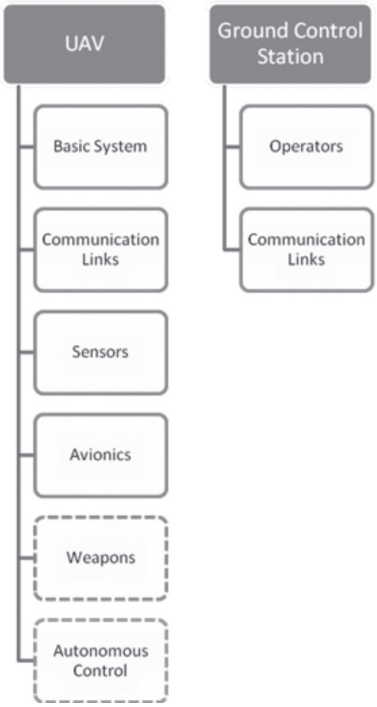
A. Preliminaries

UAVs, even at the hobby level, are increasingly complex aerial vehicles. While the majority of those available for civilian use at present are intended for short-range use under the continuous control of an operator within line of sight, autonomous UAVs with pre-programmed missions or behaviours are beginning to penetrate the civilian market, despite legal and regulatory challenges. It should be noted that the concerns stated in this paper apply equally to both of these sub-types of UAV.

UAVs are used in a variety of applications ranging from military operations such as intelligence, surveillance, target acquisition and reconnaissance (ISTAR), to civilian applications such as border control, monitoring, law enforcement, search and rescue, journalism, transportation, recreational uses, and many more. Throughout 2015-16, media reporting has routinely described new uses for UAVs where they provide significant enhancements to situational awareness or research in civilian uses; such as, to take just one example, assisting in an air accident investigation in February 2016 [2].

All of these purposes demand that UAVs are highly sensor-driven assets. It follows directly that UAVs are highly dependent on functioning sensors, and on receiving correct data from their operators and surrounding environments [3]. This dependence on real-time input, both through communication links and on-board systems, provides a wide range of vulnerabilities, following the general security guideline that any input signal to a system may be exploited to cause its malfunction [4] .

FIGURE 1: UAV COMPONENTS AND INFORMATION FLOW, FOLLOWING [3]



Before exploring past UAV incidents, a general view of UAVs from the point of view of an attacker is given in Figure 1. The UAV itself consists of a 'Basic System', being analogous to an operating system but designed to be less user-centred. This unit is connected to other components of the UAV and/or to its ground station and operator through a system of communications links, which may include any type of communication means available for interaction. A set of sensors is also available, varying according to the type of UAV. Loosely speaking, UAVs designed for purely recreational purposes are likely to have a smaller and less sophisticated set of sensors. Another entity inherent to UAVs is the avionics unit, a set of hard- and software components crucial for controlled flight. The 'autonomous' and 'weapons' systems are most likely to be found in modern military assets. While the autonomous operations system is of course security relevant in terms of vulnerability detection, the 'weapons' system is rather considered an effect carrier than a security threat. Weapons may make a UAV a more valuable target to an attacker; however, weapons are not generally considered typical targets for exploits.

For non-autonomous UAVs, the ground station and operator must also be considered. Communications links may correspond to continuous data link connections, partly-continuous connections (such as WiFi and Bluetooth, which are available upon request and within a limited range) and discrete connections which are only possible with direct access to the hardware,

such as data uploads by USB, CD-ROM, or DVD. Not considered in this article but noteworthy is that the operator himself may impose a security threat through social engineering [5].

B. Implications

In 2010, the US Federal Aviation Administration (FAA) estimated that 15,000 UAVs would be in operation in the US by 2020. In fact, by mid-2015 UAV sales there were already exceeding 15,000 each month [6]. Potentially dangerous UAV encounters by commercial airline pilots in the vicinity of airports in the US have increased accordingly. In 2014, there were 238 such reports. In 2015, the total was 650 in the first seven months [7]. It can reasonably be expected that as UAV markets develop worldwide, similar problems will be replicated elsewhere.

Users may consider that very lightweight drones cannot cause serious damage or danger, but incidents with this class of drone reported in late 2015 range from the trivial [8], through the potentially dangerous and definitely expensive [9], to the horrifying [10]. Sales predictions of up to a million small UAVs purchased for Christmas 2015 in the US raised the alarming prospect of an uncontrollable number of airborne vehicles in the hands of consumers and hobbyists with little grasp of the potential hazards of small UAV operations [11]. This prompted the FAA to rush through regulations on the use and registration of small UAVs, to be discussed below.

The explosion in UAV ownership has outstripped study of its implications, leading to a deficit of reliable studies on the actual danger, and in particular on the implications for a manned aircraft of a collision or engine strike [12]. But even within this knowledge deficit, UAV vulnerability to cyber and electronic attack stands out for an alarming degree of consumer and regulator ignorance [13]. This paper aims to assist in addressing this knowledge gap.

C. Past incidents

A series of successful cyber attacks on UAVs have been reported in recent years. Some of these were performed by researchers under laboratory conditions, while other incidents occurred ‘in the wild’. The following list is not exhaustive, and is intended only to provide evidence of the described vulnerabilities of UAVs to attack.

That UAVs may be potentially vulnerable targets in military operations has been globally acknowledged since the loss of a US RQ-170 Sentinel UAV in Iran in 2011. This incident, explored further below, called into question the US’s cyber competency, and has been frequently cited in arguments against UAV use to highlight their lack of controllability in military scenarios.

This specific incident may constitute the earliest UAV attack which led directly to public questioning of a nation’s cyber power. While the exact method by which the RQ-170 was compromised was never publicly confirmed, researchers proved subsequently that it is possible to hijack drones in flight through GPS spoofing [14]. A further relevant report was released in 2015, where members of the Critical Engineering Working Group developed a stratosphere balloon to intercept radio traffic at higher altitudes, including the frequencies used for data links between UAVs and satellites or other UAVs [15].

One line of argument suggests that this kind of attack constitutes electronic warfare (EW), rather than pure cyber attack. However, the authors of this paper consider that producing a hostile effect by introducing compromised data into an operating system meets a reasonable definition of cyber, rather than electronic, attack.

In any case, experience of current combat operations shows that the dividing lines between these different kinds of warfare are becoming increasingly blurred and irrelevant. Furthermore, regardless of the status of debate over the nature of the attack, the wide variety of available attack scenarios is one of the aspects that make UAVs especially vulnerable. From a pragmatic point of view, it does not matter how control of a software or hardware component is lost.

Besides communication links, another exploitable component is the UAVs operating system (OS) or micro-controller units as applicable. The type of OS varies between UAV manufacturers, and prototypes have been developed using smartphones as UAV control systems [16]. Thus, any known exploit in the smartphone's OS also becomes relevant in a UAV context, leading to a broader security and safety threat. It is also noteworthy that many smartphones are already compromised without the users being aware.

In 2013 Hak5 (https://hak5.org/) demonstrated a range of abuses and vulnerabilities of UAVs, including using one as a flying WiFi sniffer [17], [18]. Hak5 also reported on using a DJI Phantom 2 Vision UAV enhanced with a Pineapple WiFi and BatterPack to force Parrot AR.Drones to fly in failsafe mode, causing the AR.Drone to drop out of the sky [19]. This attack was inspired by Samy Kamkar's SkyJack Project [20] which engineers a drone 'to autonomously seek out, hack, and wirelessly take full control over any other Parrot drones [within] wireless or flying distance, creating an army of zombie drones under your control' [21]. The source code of this project is publicly available on GitHub, meaning that anybody with a rudimentary degree of skill can download it for free and run it on their own UAV.

While the examples to date have focused on interfering with data uplinks, information received from UAVs is also vulnerable to interception and exploitation. An in-combat attack intercepting the video stream between a UAV and its ground station was reported by Iraqi forces in 2009 [22]. In 2014, a research fellow student at Texas A&M University conducted a preliminary survey of the possibilities of hacking into a UAV's video stream, and the potential implications. [23]. And in February 2016 media reports based on alleged classified information stolen by Edward Snowden suggested that video feeds from Israeli UAVs had been intercepted by British signals collection installations in Cyprus [24]. But despite uncritical repetition by a wide range of media, these reports did not in fact support the suggestion of highly sophisticated decryption techniques, since the supposed intercepts were from several years earlier and of signal feeds which were unencrypted or used only basic commercial video encryption techniques [25]. Given the rapid pace of development of military UAV technology and the absence of more recent public exploits, it can be assumed that measures to prevent such simple interceptions are now in place.

3. UAVS IN THE UKRAINIAN CONFLICT

As a result of the high-profile incidents outlined above, in particular the interception of video feeds from a US UAV in Iraq [26] and the diversion and downing of another US UAV in Iran [27], the risks of military UAVs being subjected to electronic attack are well recognised. Protection of military UAV assets rightly focuses on defence against sophisticated cyber penetration or electronic attack, including data link intercepts and navigational spoofing. Offensive activity to counter adversary drone operations presumes a requirement for high-end electronic attack systems. Security in this field is in ongoing development: a US programme known as High-Assurance Cyber Military Systems (HACMS) aims to build cyber resilience for a wide range of applications including UAVs, specifically 'to create technology for the construction of high-assurance cyber-physical systems' [28].

But combat operations in eastern Ukraine in 2014-16 introduced an entirely new dimension to UAV and counter-UAV operations. In addition to drones with military-grade standards of electronic defence and encryption, a large number of civilian or amateur UAVs are in operation in the conflict. Both the Russian-backed separatists and the Ukrainian Armed Forces (VSU) have attempted to introduce UAV capabilities by using commercial civilian or home-built drones with varying degrees of modification [29].

UAVs are seeing extensive use in combat in a number of current conflicts, including in Syria, Iraq, Libya and Yemen, but the Ukrainian conflict represents the most significant use of UAVs in warfare by two opposing sides that has been documented to date. Actual operations there, in combination with studies of potential criminal or terrorist use of UAV technologies [30], provide indicators for a range of aspects of UAV use in future warfare. In addition, due in part to the vulnerabilities described in this paper, they also provide a case study of the interfaces between cyber, electronic warfare, and kinetic responses. According to one analysis, combat operations in Eastern Ukraine are 'a living lesson in how quickly war changes technology, and vice versa' [31].

These developments are being closely observed by major military UAV users. In the US view, eastern Ukraine presents 'an emerging laboratory for future 21st-century warfare' [32]. NATO too has emphasised the importance for future warfighting capability both of unmanned systems and of retaining freedom of action in the electromagnetic spectrum despite adversary capabilities [33]. It is in this respect that cyber or electronic attack on UAVs may constitute one of the most direct and immediate ways of implementing cyber power to achieve an immediate real-world effect. Close observers of Russian operations in Ukraine have noted that this effect is brought about through 'not just cyber, not just electronic warfare, not just intelligence, but […] really effective integration of all these capabilities with kinetic measures' [34].

A. Civilian UAVs

In Ukraine as elsewhere, among a wide range of uses for enhancing situational awareness on the battlefield, obtaining real-time imagery with UAVs greatly improves the accuracy and effectiveness of missile and artillery attacks. The advances in artillery effectiveness are similar

to those brought about by the use of spotters in balloons and then aircraft in the late 19th and early 20th centuries. But the irony is that in the highly sophisticated electronic warfare environment of eastern Ukraine, some of the most effective capability increments for the Ukrainian forces have been relatively inexpensive off-the-shelf consumer drones.

At the beginning of the conflict, Ukraine's military UAV stocks were mostly limited to 1970s-era Tupolev designs, limited in capability, expensive to operate, and vulnerable to attack from ground and air [35]. A number of these were indeed shot down or otherwise destroyed, although much early reporting of UAV use in the Ukraine conflict, particularly referring to US drones, was in fact disinformation [36]. In response, during 2014 programmes like the Aerorozvidka (air reconnaissance) project began to crowdfund the acquisition of UAVs for the volunteer units augmenting the VSU [37].

Many of these were off-the-shelf DJI Phantom UAVs modified for extended range and to carry Sony A-7 video cameras. The cost of acquiring and modifying each UAV was reported as being about $2,300, and the time expended in modifying and testing them followed by user training as less than a week. This compares with a reported cost of approximately $250,000 for a complete implementation package for a comparable military UAV, the US RQ-11 Raven, of which the cost of the UAV itself is approximately $30,000 [37].

However, civilian UAVs have much lower standards of protection against hostile actions. These commercial and 'entertainment' drones generally do not have intrusion detection or security mechanisms present or activated, and can be far more easily hijacked or disrupted. Unless pre-programmed for autonomous operations, small UAVs are unable to fly stealthily. Their data links, as well as being vulnerable to jamming and to cyber attacks seeking to compromise the data in order to control the UAV, broadcast continuous electromagnetic signatures that enable their detection, location and classification, as well as giving away the location of the operators. Ukrainian UAV operators have suffered casualties after being located by Russian communications intelligence operators, and targeted for mortar fire. Precautions now include frequent relocation, positioning antennas remotely, and operating from within cover [31]. The operators swiftly learned that launching from the same location more than once 'guaranteed' mortar or sniper attack [38].

Russian countermeasures thus rapidly neutralised the tactical advantage gained by Ukraine's modified civilian UAVs in late 2014. Electronic attack took the form both of GPS spoofing (feeding the UAV false information on the frequencies used to acquire satellite location data), and straightforward white noise broadcasting on the UAV's control frequency in an attempt to crash it [39]. Russia and the Russian-backed militias made use of their access to highly sophisticated and effective electronic attack technology [31]. The mismatch of resources between high-end Russian military technology and Ukrainian off-the-shelf stopgaps is stark: according to one Ukrainian UAV expert, 'They [Russia] have $7 million systems to jam drones that cost thousands of dollars' [38].

Among further planned modifications, Ukrainian software engineers began working on capability suites to militarise UAV functions, including get-you-home navigation systems for

use when GPS signals are jammed [38]. Faced with potential Russian UAV air supremacy, Ukrainian forces also requested assistance from abroad in the form of electronic countermeasures (ECM) equipment to neutralise Russian UAVs in response [40]. Monitors for the Organisation for Security and Cooperation in Europe (OSCE) also found their Schiebel S-100 Camcopter UAVs targeted by Russian-backed separatists. Attempts to shoot the OSCE UAVs down with gunfire and missiles were largely ineffective, but electronic attack including GPS jamming and spoofing caused far more serious disruption to operations, including grounding the entire OSCE UAV fleet in November 2014 [35].

There are a range of implications for NATO and other nations from UAV operations in eastern Ukraine. One clear development is that airspace for UAV operations is becoming highly contested, with air superiority considerations extending to drone operations [41]. NATO nations in particular have been led to question their long-held presumption of complete control of the air in conflict. In Ukraine, Russia employs 'tiered, multileveled [unmanned aerial systems] of all types' for reconnaissance and targeting – an entirely new challenge for NATO ground forces to deal with [42].

Other US sources note a clear distinction between Russian and US drone use. Whereas the American approach is to undertake prolonged surveillance punctuated by occasional precision strikes, the appearance of Russian UAVs is swiftly followed by intense artillery bombardment. According to Ukrainian troops, 'when they see certain types of UAVs, they know in the next 10-15 minutes, there're going to be rockets landing on top of them' [41]. In addition, Ukrainian reports suggest that Russian UAVs have operated in pairs: one at low level to draw fire, and another higher UAV to observe and provide targeting information on the Ukrainian position doing the shooting.

The UAV campaign in Ukraine has highlighted the display and use of much enhanced electronic warfare capabilities, including not only provision of false GPS data but also a range of other means of electronic attack (EA) [43]. Unofficial reports suggest some of these have already been directed from Russia at US and NATO military units visiting border regions of the Baltic states. If the Russian approach of utilising high-end EW equipment against UAVs is copied, this implies further costly investment in EA equipment which is currently available only in negligible amounts in NATO inventories.

Further afield, lessons from Ukraine can be applicable to any aspect of UAV use. All UAVs combine an array of communications systems and software, each presenting their own vulnerability to attack. These include GPS for location and height determination, digital accelerometers, camera and video suites, data processing and transmission through a variety of channels, flight control, stabilisation, autopilot capability and – for more sophisticated UAVs – pre-programmed semiautonomous operation or mission execution. All of these offer a means by which safe operation can be compromised [44]. Even geofencing software intended to prevent UAVs entering controlled or sensitive airspace, such as that developed by major drone manufacturers DJI, presents vulnerabilities [45]. Owners or adversaries may choose not to install or to disable this software, or to interfere maliciously with code or updates.

4. COUNTERMEASURES

It can be seen that many attacks on UAVs are possible due to a lack of security measures normal in other areas of IT, such as encrypted communication channels, protected software and so on. These technologies have simply not been implemented in the UAV context due to a deficient assessment of UAVs' potential as cyber-attack targets. In this section we will explore some of the ongoing efforts to establish security measures to ensure safer operation of UAVs.

A. Legislative and regulatory initiatives

While the US is undoubtedly the nation with by far the largest number of UAVs in use, it may not be the most advanced in terms of developing regulations for their use. Critiques of the legal position of drone operations highlight the central role of 'a set of rules created 70 years ago based on a chicken farm', a reference to a landmark US legal case in 1946 which determined, based on a unique set of circumstances, that the property of all landowners in the United States extends 83 feet into the air. The ruling remains in effect today [46]. Meanwhile, case law appears to be developing as a result of drones simply being shot down [47].

The significant point for the purpose of this paper is that the FAA regulations on the use and registration of small UAVs, hurriedly introduced at the end of 2015, also indicate the threat perception among US regulatory authorities. In the 211 pages of these regulations and associated commentary, cyber or electronic attack is mentioned only once, in a security proposal from the public that was not implemented. The absence of any response or commentary from the FAA suggests that this aspect of UAV hazard, and the related problem of data compromise through software or signal attack, is not being actively considered in civil operations in the US [48].

It should be emphasised that these are very different regulatory standards than those being developed for professional or military drone operations at higher altitudes and in controlled airspace where, among a range of other measures, standard air transport collision avoidance avionics are to be employed. These include active surveillance and Automatic Dependent Surveillance-Broadcast (ADS-B) to detect aircraft with transponders, TCAS 2 collision-avoidance systems, and on-board radar to detect other aircraft and validate ADS-B [49].

An informed critique of the FAA regulations suggests that a lack of threat perception is was not the only problem with them. It claims the 'interim final rule' has 'lost track of reality, claimed authority it doesn't have, and introduced rules that are destined to fail miserably [...] it is completely unworkable and the moment the FAA tries to enforce the rule, there will be hell to pay' [50].

A similar attitude appears to be held in the UK. At a public discussion in September 2015, representatives of both UK and US air traffic control authorities said that misuse of UAVs ought rightly to be a police issue, but they had been unable to raise police interest in the problem. Since this misuse is not currently a crime, no action can be taken, and consequently there is *de facto* no official concern over malicious use. A representative of the UK's largest airport company, which could expect to be directly affected by UAV misuse, said explicitly that their

concern is with accidental rather than malicious misuse. All representatives confirmed that there had been no consideration of the possibility of UAVs being hacked or hijacked through cyber compromise. The common presumption was that size mattered, but that 'bigger drones equals more risk' – the opposite of the problem in conflict situations [51].

B. Technical countermeasures

A wide range of counter-UAV technologies is rapidly becoming commercially available [52]. The most direct approach to dealing with a hostile UAV remains attempting to shoot it down; but smaller UAVs make exceptionally hard targets, and the problem of collateral damage and of where large amounts of expensive ammunition fired into the air eventually land often makes this approach prohibitive. In addition, as noted above in the context of Ukraine, firing on a UAV immediately reveals your position to other, possibly undetected, surveillance assets and invites counter-fire.

Laser weapons under development avoid the problem of collateral damage, and to some extent detection, but are limited by power consumption and disrupted by dust or fog [53]. Other inventive solutions cover a broad spectrum: 'From the Toyko [sic] police testing a net-deploying UAS to catch drones in flight to the Netherlands police training eagles to snatch quadcopters in midair, the inventiveness of the unmanned aircraft industry is evident in the counter-UAS market' [54]. Nevertheless, at the time of writing, the most promising methods for neutralising UAVs lie in cyber or electronic attack.

Blighter Surveillance Systems, Chess Dynamics, and Enterprise Control Systems of the UK have integrated radar detection, electrooptical and infrared tracking, and radiofrequency jamming to develop countermeasures for small UAVs, evaluated by the US Army in late 2015 [55]. Equipment for detecting and neutralising small UAVs has also been developed by Elta Systems, a subsidiary of Israel Aerospace Industries (IAI). Once a hostile UAV is detected, the systems use electronic attack to shut it down 'by disrupting its command link, navigation system, position location, or situational awareness' [52]. The highly portable Battelle DroneDefender makes use of directed electronic attack to jam GPS and other radio signals to a drone, causing it to land or hover without necessarily destroying it [56]. In many jurisdictions, radio frequency jamming of this kind is illegal; at the time of writing, Battelle has suspended sales and publicity 'while we evaluate the permissible applications of the product under current legislation' [57]. Similarly, recommendations that police forces should be funded for radio frequency jammers and GPS jammers to counter UAVs have to contend with the fact that their use in most countries is currently illegal, for entirely valid safety reasons [58].

In any cyber incident, whether UAV related or not, the question of attribution is fundamental. The issue of attribution in cyberspace is one that has long tormented planners and policymakers, while providing ongoing employment for lawyers. In the context of UAV control, attribution takes on a whole new dimension.

In UAV-related incidents, attribution of who is carrying out an attack is further complicated by a lack of static connections or (usually) of any logging capabilities at the UAV. Where a

UAV itself is used to carry out a physical attack, the attribution of the aggressor is even further complicated by three factors:

- The attacking UAV may not be identified at all (no ID associated with the drone, no logs available);
- The attacking UAV may be identifiable based on hardware components, but cannot be attributed to the person operating it (ID available but not registered, obsolete registration or hijacked UAV, logs only partially accessible or manipulated); and
- The human operator may be identifiable but claim not to have been in control of the drone, which at present is very difficult to prove.

Technologies to counteract these issues are available in other contexts, and their transfer to UAV operations is now under discussion. In September 2015 the EU Parliament considered a resolution to establish the 'safe use of remotely piloted aircraft systems (RPAS), commonly known as unmanned aerial vehicles (UAVs), in the field of civil aviation' [59]. This document calls for means of identifying all UAVs without regard to their size. The document implicitly acknowledges the threat recreational and civilian drones may pose to the public. It explicitly states the need for the installation of 'ID chips' in all UAVs. Furthermore, it recommends compulsory registration for buyers of UAVs throughout the EU.

As noted above, UAV registration has also been announced by the FAA in the United States [60]. A range of technical proposals for practical registration schemes is available [61], but under US regulations, tracking is carried out not by an on-board ID chip, but with a certificate and registration number to be placed on the UAV itself, in the same manner as conventional aircraft. This startlingly unimaginative approach fails to enable electronic enforcement and control methods, and is entirely unhelpful for the installation of automated logging mechanisms.

Another route to easier attribution is under consideration in the UK, which is studying the feasibility of a UAV traffic system where UAV pilots operating below 500 feet are requested to register their routes in an online database to allow tracking and avoid collisions. This attempt raises several questions:

- Usability is questionable as operators are requested to manually enter details of every flight and the exact route taken. Especially in recreational uses, this appears impractical;
- The database itself may present an additional vulnerability, as it is intended to be permanently accessible and easily updated by the public. This opens possibilities for more traditional and unsophisticated cyber attacks against the online database, such as DDoS;
- It raises questions of how the routes entered are to be monitored for correctness and accuracy, and violations addressed;
- It is unclear how false data inserted into the database are to be identified and eradicated; and
- Without a registration system, it is unclear how is the UAV is to be described uniquely within the system.

5. CONCLUSION AND OUTLOOK

Development of UAV operations continues at a startlingly rapid pace. At the time of writing, the following five scenarios belong in the future; but it is entirely possible that one or more of them will already have taken place by the time this paper reaches publication.

- At present, unmanned aircraft operations still assume permissive airspace. No UAVs have yet been announced, even by military programmes, which are able to survive in contested or denied airspace [62]. But some Ukrainian programmes to modify civilian UAVs include plans to fit weapons to them in order to target adversary drones. If this were achieved, it would be the first documented case of UAV-on-UAV warfare, and akin to the very earliest days of aerial combat during the First World War when pilots of unarmed reconnaissance aircraft began to take rifles in the cockpit to take potshots at each other.
- Ukrainian forces repeatedly refer to being 'swarmed' by Russian drones. But in the US, two separate programmes are testing genuine swarming capabilities, where large number of autonomous UAVs act as a mass to overwhelm an adversary's defences. DARPA's Gremlins programme is to trial launching numbers of small UAVs from aircraft to carry out coordinated and distributed operations. The Office of Naval Research's Locust (LowCost UAV Swarming Technology) programme, understandably, envisages launching the swarm of small UAVs from ships [63].
- Mixing manned and unmanned operations will also be on trial. In an approach with some similarities to the concept for manned aircraft with different roles to cooperate and share information using the Talon HATE pod [64], the US Air Force Research Laboratory's 'Loyal Wingman' programme aims to team manned fighters with 'survivable UAVs that collaborate as offboard sensors, jammers and weapons trucks' [65].
- Progress in regulating civilian UAV use is likely to lead to additional sensors and communications devices to avoid restricted airspace and collisions with other aircraft. Autonomous systems small enough to be mounted on micro-UAVs, including implementations of the ADS-B system used on manned aircraft, are already available [66]. Secure failsafe mechanisms can be expected to be built in to UAVs as standard, providing for controlled descent or return to base when ground or GPS communications are lost or when the UAV detects electronic attack. But systems such as these present yet another vulnerability to hostile interference: if their software, data or communications are not adequately protected, then they too are open to cyber or electronic attack.
- Potential future deliberate use of UAVs for terrorist purposes remains a hot topic. In early 2016, a report highlighting the risks led to alarmist headlines in the US and Europe [67], [68]. But even this report focused exclusively on the prospects of terrorist organisations developing their own UAVs, and did not address the potential for cyber hijacking of third party UAVs in order to carry out attacks.

In summary, the rapid development of UAV capabilities is far outstripping concepts and

procedures for ensuring their security. It is commonly repeated that the challenge of ensuring cyber security overall arises largely from the fact that the internet was designed to be fundamentally insecure. By contrast, the current state of development of UAVs presents an opportunity to recognise the problems outlined in this paper, and consequently begin to build in protection against cyber attack as standard.

REFERENCES

[1] Dr Shima Keene. (2015, December) 'Lethal and Legal? The Ethics of Drone Strikes', *U.S. Army War College - Strategic Studies Institute*. [Online]. http://www.strategicstudiesinstitute.army.mil/pubs/display.cfm?pubID=1304.

[2] John Croft. (2016, February) 'Drone Aids TransAsia Flight 222 Accident Investigation', *Aviation Week*. [Online]. http://aviationweek.com/commercial-aviation/drone-aids-transasia-flight-222-accident-investigation.

[3] Kim Hartmann and Christoph Steup, 'The Vulnerability of UAVs to Cyberattacks', in *5th International Conference on Cyber Conflict*, Tallinn, Estonia, 2013.

[4] Matt Bishop, 'Introduction to Computer Security'. Boston, USA: Addison-Wesley, 2004.

[5] Kevin Mitnick, 'The Art of Deception: Controlling the Human Element of Security': John Wiley & Sons, 2003.

[6] Aaron Karp. (2015, October) 'Congress to hold UAV safety hearing Oct. 7', *ATWonline.com*. [Online]. http://atwonline.com/government-affairs/congress-hold-uav-safety-hearing-oct-7.

[7] John Croft. (2015, October) 'DOT: Register Your Drones Or Face FAA Penalties', *Aviation Week*. [Online]. http://aviationweek.com/commercial-aviation/dot-register-your-drones-or-face-faa- penalties.

[8] Cyrus Farivar. (2015, November) 'Drone collides with Seattle Ferris wheel, busts through plastic table', *Ars Technica*. [Online]. http://arstechnica.com/tech-policy/2015/11/drone-collides-with-seattle-ferris-wheel-busts-through-plastic-table/.

[9] Megan Guess. (2015, June) 'Drone flying over forest fire diverts planes, costs US Forest Service $10K', *Ars Technica*.

[10] Cyrus Farivar. (2015, December) 'Toddler loses eyeball after errant drone slices it in half', *Ars Technica*. [Online]. http://arstechnica.com/tech-policy/2015/12/toddler-loses-eyeball-after-errant-drone-slices-it-in-half/.

[11] Aaron Karp. (2015, September) 'FAA Nightmare: A Million Christmas Drones', *Aviation Week*. [Online]. http://aviationweek.com/commercial-aviation/faa-nightmare-million-christmas-drones.

[12] Aviation Week & Space Technology. (2015, September) 'Editorial: Get Data On Risk UAS Pose To Air Traffic', *Aviation Week*. [Online]. http://aviationweek.com/commercial-aviation/editorial-get-data-risk-uas-pose-air-traffic.

[13] Thomas Fox-Brewster. (2016, March) 'Police Drone Can Be Commandeered From Over A Mile Away, Hacker Claims', *Forbes.com*. [Online]. http://www.forbes.com/sites/thomasbrewster/2016/03/02/surveillance-drone-hacked/.

[14] Andrew J. Kerns, Daniel P. Shepard, Jahshan A. Bhatti, and Todd E. Humphreys, 'Unmanned Aircraft Capture and Control Via GPS Spoofing', *Journal of Field Robotics*, vol. 31, no. 4, 2014.

[15] Peter König. (2015, September) 'Hacker starten Stratosphärenballon, um Drohnen-Funk mitzuschneiden' ('Hackers use stratosphere balloon to intercept UAV radio traffic'), *Heise.de*. [Online]. http://www.heise.de/make/meldung/Hacker-starten-Stratosphaerenballon-um-Drohnen-Funk-mitzuschneiden-2823100.html.

[16] Heise Online. (2015, November) 'PhoneDrone Ethos: Drohne nutzt Smartphone als Steuerungsrechner' ('PhoneDrone Ethos: Drone uses Smartphones as Controlunit'), *Heise.de*. [Online]. http://www.heise.de/newsticker/meldung/PhoneDrone-Ethos-Drohne-nutzt-Smartphone-als-Steuerungsrechner-2912422.html.

[17] Hak5. (2014, January) 'Pineapple Drone, Rooftop Packet Sniffing And Offline Archival Backup', *hak5.org*. [Online]. http://hak5.org/episodes/hak5-1520.

[18] Ricky Hill. (2013, March) 'Phantom Network Surveillance UAV / Drone - Defcon', *Defcon.org*. [Online]. https://www.defcon.org/images/defcon-21/dc-21-presentations/Hill/DEFCON-21-Ricky-Hill-Phantom-Drone-Updated.pdf.

[19] Hak5. (2013, December) 'Drones Hacking Drones', *hak5.org*. [Online]. http://hak5.org/episodes/hak5-1518.

[20] Kamkar, Samy. (2013, December) 'SkyJack: autonomous drone hacking'. [Online]. http://samy.pl/skyjack/.

[21] Kamkar, Samy. (2013, December) 'SkyJack', *Github.com*. [Online]. https://github.com/samyk/skyjack.
[22] BBC News. (2009, December) 'Iraq insurgents 'hack into video feeds from US drones'', *BBC.co.uk*. [Online]. http://news.bbc.co.uk/2/hi/middle_east/8419147.stm.
[23] Emy Rivera, Robert Baykov, and Goufei Gu, 'A Study On Unmanned Vehicles and Cyber Security', Texas, USA, 2014.
[24] Dan Lamothe. (2016, January) 'U.S. and Britain hacked into feeds from Israeli drones and fighter jets, according to report', *Washington Post*. [Online]. https://www.washingtonpost.com/news/checkpoint/wp/2016/01/29/u-s-and-britain-hacked-into-feeds-from-israeli-drones-and-fighter-jets-according-to-report/.
[25] Samvartaka blog. (2016, February) 'Cryptanalysis of intercepted Israeli drone feeds', *Github.io*. [Online]. http://samvartaka.github.io/cryptanalysis/2016/02/02/videocrypt-uavs.
[26] Mike Mount and Elaine Quijano. (2009, December) 'Iraqi insurgents hacked Predator drone feeds U.S. official indicates', *CNN.com*. [Online]. http://edition.cnn.com/2009/US/12/17/drone.video.hacked/.
[27] Scott Peterson. (2011, December) 'Exclusive: Iran hijacked US drone, says Iranian engineer', *Christian Science Monitor*. [Online]. http://www.csmonitor.com/World/Middle-East/2011/1215/Exclusive-Iran-hijacked-US-drone-says-Iranian-engineer-Video.
[28] Raymond Richards. (Undated) 'High-Assurance Cyber Military Systems (HACMS)', *DARPA.mil*. [Online]. http://www.darpa.mil/program/high-assurance-cyber-military-systems.
[29] Joe Gould. (2015, August) 'Electronic Warfare: What US Army Can Learn From Ukraine', *Defense News*. [Online]. http://www.defensenews.com/story/defense/policy-budget/warfare/.um=email&utm_term=%2ASituation%20Report&utm_campaign=SitRep0803.
[30] Dr. Robert J. Bunker. (2015, August) 'Terrorist and Insurgent Unmanned Aerial Vehicles: Use, Potentials, and Military Implications', *U.S. Army War College - Strategic Studies Institute*. [Online]. http://www.strategicstudiesinstitute.army.mil/pubs/display.cfm?pubID=1287.
[31] Patrick Tucker. (2015, March) 'In Ukraine, Tomorrow's Drone War Is Alive Today', *Defence One*. [Online]. http://www.defenseone.com/technology/2015/03/ukraine-tomorrows-drone-war-alive-today/107085/.
[32] Graham Warwick. (2016, January) 'Assisting The Human Central to Pentagon's Third Offset', *Aviation Week*. [Online]. http://aviationweek.com/defense/assisting-human-central-pentagon-s-third-offset.
[33] North Atlantic Treaty Organisation. (2015, August) 'Framework for Future Alliance Operations', *NATO.int*. [Online]. http://www.act.nato.int/images/stories/media/doclibrary/ffao-2015.pdf.
[34] Sydney J. Freedberg. (2015, November) 'Army Fights Culture Gap Between Cyber & Ops: "Dolphin Speak"', *BreakingDefense.com*. [Online]. http://breakingdefense.com/2015/11/army-fights-culture-gap-between-cyber-ops-dolphin-speak/.
[35] Adam Rawnsley. (2015, February) 'War is Boring', *Medium.com*. [Online]. https://medium.com/war-is-boring/ukraine-scrambles-for-uavs-but-russian-drones-own-the-skies-74f5007183a2.
[36] Maksym Bugriy. (2014, June) 'The Rise of Drones in Eurasia (Part One: Ukraine)', *JamesTown.org*. [Online]. http://www.jamestown.org/single/?tx_ttnews%5Btt_news%5D=42536.
[37] Nolan Peterson. (2015, March) 'Ukraine's Grassroots Drone Program Takes Flight', *The Daily Signal*. [Online]. http://dailysignal.com/2015/03/12/ukraines-grassroots-drone-program-takes-flight/.
[38] Christian Borys. (2015, April) 'Crowdfunding a war: Ukraine's DIY drone-makers', *The Guardian*. [Online]. http://www.theguardian.com/technology/2015/apr/24/crowdfunding-war-ukraines-diy-drone-makers.
[39] Nicholas Lazaredes. (2015, April) 'Ukraine's DIY drone war: Self-taught soldiers facing up to Russian-backed war machine', *ABC.net*. [Online]. http://www.abc.net.au/news/2015-04-22/ukraines-diy-drone-war/6401688.
[40] Patrick Tucker. (2015, February) 'How US Technology Could Help Ukraine Without 'Arming' It', *Defense One*. [Online]. http://www.defenseone.com/technology/2015/02/how-us-technology-could-help-ukraine-without-arming-it/104931/.
[41] Sydney J. Freedberg. (2015, October) 'Russian Drone Threat: Army Seeks Ukraine Lessons', *BreakingDefense.com*. [Online]. http://breakingdefense.com/2015/10/russian-drone-threat-army-seeks-ukraine-lessons/.
[42] Andrew Tilghman. (2015, December) 'Advanced Russian air power, jammers are focus of U.S. troops', *Military Times*. [Online]. http://www.militarytimes.com/story/military/pentagon/2015/12/10/advanced-russian-air-power-jammers-focus-us-troops/77090544/.
[43] SC Magazine. (2015, October) 'Russia overtaking US in cyber-warfare capabilities', *SCMagazine.com*. [Online]. http://www.scmagazine.com/russia-overtaking-us-in-cyber-warfare-capabilities/article/450518/.
[44] David Esler. (2015, September) 'What A Business Aviation Flight Department Needs To Know About UAVs', *Aviation Week*. [Online]. http://aviationweek.com/print/business-aviation/what-business-aviation-flight-department-needs-know-about-uavs.

[45] Emily Reynolds. (2015, November) 'DJI update enforces drone no-fly zones across Europe and USA', *Wired*.
[46] Kieren McCarthy. (2016, January) 'Bloke sues dad who shot down his drone – and why it may decide who owns the skies', *The Register*. [Online]. http://www.theregister.co.uk/2016/01/07/drone_lawsuit_who_owns_the_skies/.
[47] Cyrus Farivar. (2015, October) '"Drone Slayer" cleared of charges: "I wish this had never happened"', *Ars Technica*. [Online]. http://arstechnica.com/tech-policy/2015/10/drone-slayer-cleared-of-charges-i-wish-this-had-never-happened/.
[48] Federal Aviation Administration. (2015, December) 'Registration and Marking Requirements for Small Unmanned Aircraft', *FAA.gov*. [Online]. https://www.faa.gov/news/updates/media/20151213_IFR.pdf.
[49] Graham Warwick. (2015, October) 'First Interim Standards For Unmanned Aircraft Unveiled', *Aviation Week*. [Online]. http://aviationweek.com/commercial-aviation/first-interim-standards-unmanned-aircraft-unveiled.
[50] Kieren McCarthy. (2015, December) 'FAA introduces unworkable drone registration rules in time for Christmas', *The Register*. [Online]. http://www.theregister.co.uk/2015/12/14/faa_drone_registration_rules/.
[51] Chatham House. (2015, September) 'Dealing with Drones: A Look at the Regulatory Challenges of Remotely Piloted Aircraft Systems', *Chatham House Seminar*. [Online]. https://www.chathamhouse.org/event/dealing-drones-look-regulatory-challenges-remotely-piloted-aircraft-systems.
[52] David Eshel and John M. Doyle. (2015, November) 'UAV Killers Gain Role Against Growing Threat', *Aviation Week*. [Online]. http://aviationweek.com/defense/uav-killers-gain-role-against-growing-threat.
[53] Daniel Culpan. (2015, August) 'Boeing's latest drone destroyer is the stuff of nightmares', *Wired*.
[54] Graham Warwick. (2016, February) 'Counter-UAS Special Report: The Countermeasures Options', *Aviation Week*. [Online]. http://aviationweek.com/technology/counter-uas-special-report-countermeasures-options.
[55] Graham Warwick. (2015, December) 'Countering Unmanned Defense & Space Technologies To Watch In 2016', *Aviation Week*. [Online]. http://aviationweek.com/defense/defense-space-technologies-watch-2016-0.
[56] Swati Khandelwal. (2015, October) 'First Ever Anti-Drone Weapon that Shoots Down UAVs with Radio Waves', *The Hacker News*. [Online]. http://thehackernews.com/2015/10/drone-defender-gun.html.
[57] Battelle. (2016, April) 'Battelle DroneDefender', *Battelle.org*. [Online]. http://www.battelle.org/our-work/national-security/tactical-systems/battelle-dronedefender.
[58] Kieren McCarthy. (2016, January), 'Beware the terrorist drones! For they are coming! Pass new laws!', *The Register*. [Online]. http://www.theregister.co.uk/2016/01/11/beware_terrorist_drones/.
[59] Jacquelin Foster. (2015, September) 'Report on the safe use of remotely piloted aircraft systems (RPAS), commonly known as unmanned aerial vehicles (UAVs), in the field of civil aviation', *EUROPA.eu*. [Online]. http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2015-0261+0+DOC+XML+V0//EN.
[60] Federal Aviation Administration. (2015, December) 'Press Release – FAA Announces Small UAS Registration Rule', *FAA.gov*. [Online]. http://www.faa.gov/news/press_releases/news_story.cfm?newsId=19856.
[61] Jared Ablon, Steve Crocker, Benjamin D. Marcus, and Gregory S. McNeal. (2016, February) 'Robust and Scalable UAS Registration: Key Technology Issues And Recommendations', *SUASNews.com*. [Online]. www.suasnews.com/wp-content/uploads/2016/02/AirMap_White-Paper_UAS-Registration_02042016.pdf.
[62] Graham Warwick and Larry Dickerson. (2015, December) 'Military UAVs Mark Time As Civil Market Advances', *Aviation Week*. [Online]. http://aviationweek.com/print/defense/military-uavs-mark-time-civil-market-advances.
[63] Graham Warwick. (2015, December) '"Swarm Theory", Defense & Space Technologies To Watch In 2016', *Aviation Week*. [Online]. http://aviationweek.com/defense/defense-space-technologies-watch-2016-0.
[64] Tyler Rogoway. (2015, December) 'Here's The First Shot Of The F-15C Pod That Will Change How The Air Force Fights', *Foxtrotalpha*. [Online]. http://foxtrotalpha.jalopnik.com/here-s-the-first-shot-of-the-f-15c-pod-that-will-change-1750314539.
[65] Graham Warwick. (2015, December) '"Team Players", Defense & Space Technologies To Watch In 2016', *Aviation Week*. [Online]. http://aviationweek.com/defense/defense-space-technologies-watch-2016-0.
[66] Graham Warwick. (2016, January) 'Tiny ADS¬B Provides UAV Sense¬-and--avoid', *Aviation Week*. [Online]. http://aviationweek.com/commercial-aviation/week-technology-jan-4-8-2016.
[67] Matt Burges, 'UK at risk from 'simple and effective' terrorist drone attacks', *Wired*, January 2016.
[68] Tony Osborne. (2015, January) 'Terror by Drone', *Aviation Week & Space Technology (print edition)*, pp. 28-29.

[69] Alan Levin. (2015, May) ‘FAA introduces unworkable drone registration rules in time for Christmas’, *Bloomberg.com*. [Online]. http://www.bloomberg.com/news/articles/2015-05-29/google-s-solar-fueled-cyber-drone-crashes-during-new-mexico-test.
[70] Christian Czosseck, ‘State Actors and their Proxies in Cyberspace’, in *Peacetime Regime for State Activities in Cyberspace*. Tallinn, Estonia: NATO Cooperative Cyber Defence Centre of Excellence, 2013.
[71] Kim Hartmann and Christoph Steup, ‘N3P: A Natural Privacy Preserving Protocol for Electronic Mail’, in *4th International Conference on Cyber Conflict*, Tallinn, Estonia, 2012.
[72] Daniel T. Kuehl, ‘From Cyberspace to Cyberpower: Defining the Problem’, in *Cyberpower and National Security*.: Potomac Books Incorporated, 2009.
[73] Heli Tiirmaa-Klaar, ‘Cyber Diplomacy: Agenda, Challenges and Mission’, in *Peacetime Regime for State Activities in Cyberspace*. Tallinn, Estonia: NATO Cooperative Cyber Defence Centre of Excellence, 2013.
[74] Eray Yagdereli, Cemal Gemci, and A. Ziya Aktas, ‘A study on cyber-security of autonomous and unmanned vehicles’, *The Journal of Defense Modelling and Simulation: Applications, Methodology, Technology*, vol. 12, no. 4, 2015.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Assessing the Impact of Aviation Security on Cyber Power

Martin Strohmeier
Department of Computer Science
University of Oxford
Oxford, United Kingdom
martin.strohmeier@cs.ox.ac.uk

Vincent Lenders
Science and Technology
armasuisse
Thun, Switzerland
vincent.lenders@armasuisse.ch

Matthias Schäfer
Department of Computer Science
University of Kaiserslautern
Kaiserslautern, Germany
schaefer@cs.uni-kl.de

Ivan Martinovic
Department of Computer Science
University of Oxford
Oxford, United Kingdom
ivan.martinovic@cs.ox.ac.uk

Matt Smith
Department of Computer Science
University of Oxford
Oxford, United Kingdom
matthew.smith@cs.ox.ac.uk

Abstract: We analyse the impact of new wireless technology threat models on cyber power, using the aviation context as an example. The ongoing move from traditional air traffic control systems such as radar and voice towards enhanced surveillance and communications systems using modern data networks causes a marked shift in the security of the aviation environment. Implemented through the European SESAR and the US American NextGen programmes, several new air traffic control and communication protocols are currently being rolled out that have been in the works for decades. Unfortunately, during their development the shifting wireless technology threat models were not taken into account. As technology related to digital avionics is getting more widely accessible, traditional electronic warfare threat models are fast becoming obsolete.

This paper defines a novel and realistic threat model based on the up-to-date capabilities of different types of threat agents and their impact on a digitalised aviation communication system. After analysing how the changing technological environment affects the security of aviation technologies, current and future, we discuss the reasons preventing the aviation industry from

quickly improving the security of its wireless protocols. Among these reasons, we identify the existing tradition of the industry, the prevalence of legacy hard- and software, major cost pressures, slow development cycles, and a narrow focus on safety (as opposed to security). Finally, we analyse how this major technological shift informs the future of cyber power and conflict in the aviation environment by looking at tangible effects for state actors.

Keywords: *aviation security, cyber power, critical infrastructures, wireless attacks, communication security, aviation privacy*

1. INTRODUCTION

Modern wireless data networks are becoming increasingly important as a communication tool for aircraft and ground surveillance alike. While it has been well known for years within the computer security community [1] that both current and future aviation communication and surveillance systems do not offer enough – or any – protection against cyber attack, the concrete impact on cyber power has not been analysed so far.

As wireless communications technology advanced rapidly over the past two decades, commercial-off-the-shelf (COTS) hardware with the ability to affect wireless systems within aviation has become widely available. The long technological upgrade cycles for digital avionics systems guarantee that the employed wireless technology becomes dated at some point during its life cycle, which profoundly affects the security of these technologies in particular. As a result, traditional electronic warfare threat models have become obsolete and can no longer provide the basis for the security analysis of civil aviation.

Instead, a modern threat model must consider the impact of potential attacks not only on the electromagnetic spectrum but also on the increasingly digitalised aviation communications system as a whole. Combining both modern cyberspace operations, and traditional electronic warfare under the umbrella of Cyber Electromagnetic Activities (CEMA) is a relatively new concept which is quickly gaining importance within the defence community [2]. This article examines how CEMA can directly affect the critical infrastructure system of aviation with potentially devastating consequences.

We analyse how the wide proliferation of software-defined radio hardware and the accompanying development and accessibility of software tools and knowledge enable a large group of actors to both passively and actively engage with the aviation communications system. Because of these advances, the technological advantage and obscurity on which aviation's communications security is based has become untenable. As proprietary knowledge on aviation protocols is widely accessible, even unsophisticated actors with few resources cannot be prevented accessing the wireless communication used for ensuring air safety any more.

As the awareness of this issue has only started to increase recently [3], newly developed future communication technologies such as the Automatic Dependent Surveillance Broadcast protocol (ADS-B) do not solve this security problem but rather exacerbate it. We postulate that these existing security and privacy issues already have a measurable impact on current cyber conflicts, and analyse how this democratisation of wireless capabilities affects the cyber power of state actors.

The contributions of this paper are:

- We analyse the impact of recent technological advances in wireless communications and the digitalisation of avionics on the security of aviation protocols. Based on these insights, we develop a novel realistic threat model replacing the traditional electronic warfare model.
- Using the newly developed threat model, we classify the relevant threat agents based on their motivation and wireless capabilities. We analyse the security of wireless air traffic control protocols, current and future, based on our taxonomy.
- Finally, we discuss the impact of the changing threat environment on the cyber power of state actors in the aviation system. We postulate a democratisation of power, shifting away from nation states towards a much wider range of actors.

The remainder of this article is organised as follows. Section 2 discusses the new threat model faced by actors in aviation, and then Section 3 examines some of the threat agents involved in this model. Section 4 provides a security analysis of exemplary current and future aviation technology. Section 5 looks at the reasons for the current lack of security within civil aviation. Section 6 discusses the impact of security- and privacy-related technology advances on nation state actors and their cyber power. Section 7 briefly presents the related work on critical infrastructures and aviation security in particular. Finally, Section 8 concludes this work.

2. THE NEW CYBER THREAT MODEL IN AVIATION

In this section, we discuss recent advances made in wireless technologies, and how they have changed the threat landscape in the aviation context. As civil aviation systems increasingly move towards modern digital data networks, we further argue that this increased digitisation and automation leads to new vulnerabilities not present in the traditional aviation safety mind-set. We illustrate the impact of these developments on the security of aviation.

A. Recent advances in wireless technology

The technological advances in wireless technology happening in the late 1990s and 2000s drastically changed the assumptions about the capabilities of adversaries in wireless communication settings. One of the main drivers of this development has been software-defined radio (SDR) technology. SDRs were first developed for military and closed commercial use in the 1990s followed by open-source projects such as GNU Radio [4], which was released in 2001. In conjunction with the availability of cheap commercial off-the-shelf SDR hardware, new

technological capabilities became available to a large group of people. Anyone with a relatively basic technological understanding can now receive, process, craft, and transmit arbitrary radio signals such as those used in aviation systems. Where previously radio hardware needed to be purpose-built, an expensive and complicated endeavour feasible only for specialists, SDRs can be programmed and seamlessly adapted using software easily available on the Internet.

B. Move towards digital communication networks and automation

Complementing the technological advances available to the general public, we observe a trend in aviation towards transmitting data using unauthenticated digital communication networks. This digital data, which is as diverse as flight clearances, aircraft positions, or passenger information, is subsequently processed by increasingly automated systems on the ground and in the aircraft, which implicitly relies on the integrity of the received information to ensure the safety of air travel. Without authentication of the underlying protocols, attacks on the data link level are inherently more difficult to detect for both aviation systems and their users than attacks on traditional analogue technologies such as voice communication or primary surveillance radar.

C. Impact on aviation security

Together, the discussed technological trends and advances have had a profound impact on the security of wireless aviation protocols and consequently caused a fundamental shift in the applicable threat model. The move towards unsecured digital networks and increased deployment of COTS hard- and software in avionics enables new adversarial groups, which stand far outside the former military-inspired electronic warfare threat model [5].

The advent of SDR technology has provided a surge of applications for radio communications in general. The former assumption that access to the frequencies used by important communication technologies is hard has been voided. Modulations of virtually all radio applications are well known and made available freely through the SDR community. Thus, the ability to eavesdrop and manipulate any communication wireless channel is available to any interested observer without the requirement for significant resources and specialist knowledge. Examples of such possibilities are the trivial access to mobile phone networks, satellite signals, television channels, or wireless sensor networks.

One of the most active and enthusiastic SDR communities is concerned with aviation communication and flight tracking. Using, for example, the popular RTL-SDRs, a $10 USB stick repurposed as a software-defined radio receiver, a plane-spotter can choose between several different software options to receive virtually all air traffic communication protocols in use today (e.g. ADS-B [6]). Countless enthusiasts and volunteers around the world use such hard- and software to power a multitude of services such as flightradar24.com, opensky-network.org, or adsbexchange.com, where an ever-increasing number of flight movements can be followed live and without delay. Data from flight trackers has been involved regularly in investigations following flight incidents such as the Germanwings crash [7], or the two Malaysian Airlines aircraft lost over the Ukraine and the Indian Ocean in 2014, illustrating the impact of the changing communications landscape on aviation.

With more powerful SDRs, which are capable of sending as well as receiving, becoming cheaper and widely available, it is possible to manipulate virtually all aspects of the wireless channel used by aviation protocols [8]. These possibilities stand in stark contrast to the pre-SDR electronic warfare threat model focused on nation states being the only actors with the expensive and sophisticated capabilities required to attack ATC systems. The impact of this development on ATC is discussed in Section 4.

3. A TAXONOMY OF CYBER THREAT AGENTS

Based on the insights from the previous section, we develop a new threat model for wireless attacks in the aviation context focusing on CEMA threats. We analyse possible attackers based mainly on: a) their resources; b) their subject-matter expertise; and c) their motivation. Table 1 presents the threat agents applicable to wireless security in aviation, which we go on to discuss in detail. Our taxonomy is very loosely inspired by the relevant NIST definitions [9], but adapted for the unique context of the cyber-physical aviation system. While our approach to threat agents in aviation is novel, we believe that tying it into the existing NIST framework leads to easier application in practice. Our taxonomy provides new insights into the specific technological capabilities of different classes of threat agents, and how these can be exploited to achieve their respective goals, even in light of potential countermeasures.

TABLE 1: OVERVIEW OF THREAT AGENTS

Threat	Resources	Type	Goal/Motivation
Passive Observers	None - Very low	Passive	Information collection / Financial or personal interest
Script Kiddies / Hobbyists	Low	Active	Any noticeable impact / Thrill and recognition
Cyber Crime	Medium - High	Active	Maximising impact / Financial gains using e.g. blackmail or valuable information
Cyber Terrorism	Low - Medium	Active	Political or religious motivation / Massive disruption and casualties
Nation State	Unlimited	Active	Weapons / Targeting specific, potentially military objects

A. Passive observers

Passive observers are interested people who exploit the open nature of air traffic communication protocols to glean information. This class of threat agents does not actively interfere with air traffic communication, but instead uses public and private websites and mobile applications, which display air traffic and its communications in real time, to gather information about private or secret air traffic movements. Alternatively, they can employ cheap SDR receivers to gather their own undistorted picture of all air traffic in their vicinity, in real time or stored for later analysis. The information collected by such merely passive observers can be exploited in many ways, ranging from privacy concerns to the detection of military operations, which are discussed in detail in Section 6.

B. Script kiddies and hobbyists

Script kiddies and hobbyists are the lowest active threat in our model, based on their abilities concerning both hardware and knowledge. Their aim is to exploit well-known security holes with existing, easy-to-use attacks with typically low sophistication. Their motivation is regularly not rational; instead any identifiable impact is sought for thrill and recognition [9]. We assume an attack to be the following:

> Using a programmable transponder, they listen in to legitimate radio communication, modify the call sign and/or information such as position and velocity, and play it back. The objective of the attacker is to have their signals either shows up as a new aircraft with an unexpected call sign, or as an existing aircraft causing conflicting information. We assume that the attacker is on the ground and sends with the standard parameters of their transponder.

Hobbyists are typically interested in plane-spotting and more familiar with the norms and protocols in modern ATC, either due to personal interest in aviation or because it relates to their job. They are also more knowledgeable about radio communication and the basic characteristics of the wireless channel. They have access to SDRs and are able to operate them with matching software frameworks such as GNU Radio. Their attack is similar to that of the script kiddies, but it is not detected by naïve plausibility checks on the data link or physical level.

C. Cyber crime

The cyber crime attacker class seeks to attack systems for monetary gain. With sufficient subject-matter knowledge, software-defined radios, and potentially even small unmanned aerial vehicles (UAV), they are able to inject new messages or modify existing ones in such ways that they are not flagged by current detection systems. Cyber crime attackers are typically interested in causing maximum damage and exerting credible threats, as a pre-requisite for blackmail or to take advantage of captured inside knowledge.[1] Consequently, they are seeking to exploit any possible and effective way to attack ATC and aircraft systems.

D. Cyber terrorism

Attacks on cyber-physical systems powering critical infrastructures such as aviation are a natural target for terrorists and politically motivated attacks. Terrorists seek to threaten national security, cause mass casualties, weaken the economy, and damage public morale and confidence [9]. By exploiting the vulnerabilities in wireless aviation communications, terrorist groups, which traditionally hijack or crash planes using physical weapons, could mount attacks on planes from the ground and from safe distances.

E. Nation states

With sufficient knowledge of intrusion detection systems and near-unlimited resources, it is possible to bypass plausibility checks and redundancy-based defences. While it becomes increasingly difficult to deceive many ATC systems at the same time, it is possible. However, we argue that it is only achievable by a nation state actor and part of the electronic warfare threat model traditionally outside the scope of securing civil aviation. In this case, non-transparent

[1] Aircraft movement information has allegedly been used for stock trading, see, e.g., [35]

countermeasures such as authentication through cryptographic means may help further, although this requires a complete overhaul of the protocol set and administrative planning [10]. Thus, it is unlikely to happen in the near future.

4. THE CASE OF AIR TRAFFIC CONTROL SURVEILLANCE

In order to demonstrate more clearly how aviation has to deal with the changing cyber security threat, this section presents an example technology: air traffic surveillance. The set of technologies used is integral to the safe operation of airspace, yet as it becomes more technologically advanced, it also becomes more insecure. This change is representative of avionic technology as a whole [11]. Throughout this section, we assess the technologies with respect to our threat model as seen in Table 2. From this, we attempt to match which systems are 'in reach' of a given attacker, which is summarised in Table 3.

TABLE 2: SUMMARY OF SURVEILLANCE TECHNOLOGIES

Technology	Ground/Air Dependent	Cost[2]	Deployment Status
PSR	Ground	High	In use
SSR	Ground	High	In use
TCAS (STANDARD)	Air	Moderate	Mandate by 2015[3]
TCAS (HYBRID)	Air	Moderate	Optional
ADS-B	Air	Low	Mandate by 2020
WAM	Ground	High	In deployment

A. Surveillance fundamentals

In order for ATC to safely manage airspace, each controller needs to understand the status of each aircraft under their control. Traditionally, Primary Surveillance Radar (PSR) and Secondary Surveillance Radar (SSR) in various forms have fulfilled this role since World War II.

Both systems were designed at a time when radio transmission required a great financial investment and expertise. Hence, no thought was given to securing the systems, as it was presumed that they would remain out of reach. The rise of SDRs voided this assumption; they marked the shift from potential attackers being well resourced to those with much less resource and capability.

PSR describes non-cooperative radar localisation systems. In civil aviation, these typically employ a rotating antenna radiating a pulse position-modulated and highly directional electromagnetic beam on a low GHz band. Potential targets in the airspace reflect the pulses; measurement of the bearing and round trip time of these reflections provides the target's

2 High cost is considered to be >$1 million, moderate > $100,000, low < $100,000.
3 For most civil aircraft, see Section 4.B.2.

position. Whilst PSR is not data-rich, it is relatively hard to attack as it relies on physical properties [11]. As such, we consider attacks on PSR to be out of scope of all but sophisticated nation state actors.

SSR is a cooperative technology with modern versions including the so-called transponder Modes A, C, and S. SSR provides more target information on ATC radar screens compared to PSR. Ground stations interrogate aircraft transponders using digital messages on the 1030 MHz frequency, which reply with the desired information on the 1090 MHz channel. Commodity hardware can receive and transmit on these frequencies, making them accessible to attack [3]. Very few skills are needed to receive SSR today, bringing it into reach of script kiddies and hobbyists, who might also be able to disturb ATC systems by simply injecting or replaying old SSR messages. To mount more sophisticated active attacks such as denial of service, slightly more skill and resource are needed, as is a definite motivation to disrupt, placing it in the cyber terrorism and cyber crime domains.

B. Current and next generation surveillance

NextGen and SESAR incorporate a range of surveillance technologies as part of the effort to reduce costs and increase efficiency [12]. Even though these technologies are in the early stages of deployment, they were designed decades ago. The result is that these systems have yet to be adapted to a modern threat model.

Mode S is a particularly important part of the current SSR system. It provides two systems of increasing significance in modern aviation: Automatic Dependent Surveillance-Broadcast (ADS-B), and Traffic Collision and Avoidance System (TCAS).[4] These systems are being rolled out as key factors in surveillance, in conjunction with multilateration techniques to provide redundancy. Due to an intentional lack of confidentiality, all SSR systems are subject to eavesdropping attacks by passive observers.

1. ADS-B

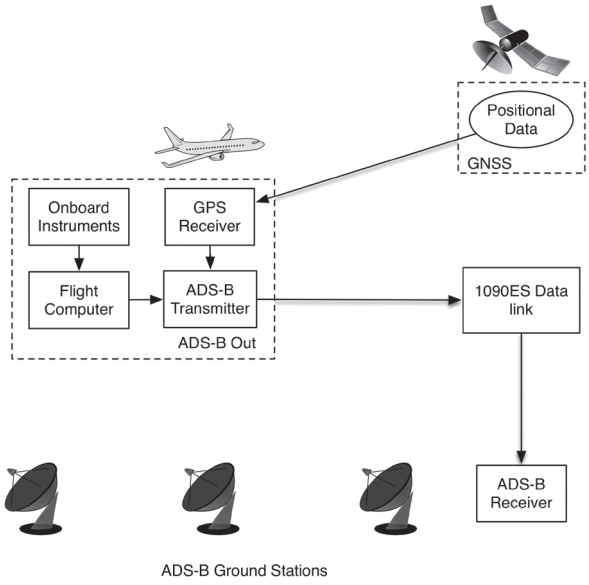
ADS-B is a protocol in which aircraft continually broadcast their own ID, position and velocity as well as further information such as intent or urgency codes. These broadcasts do not require interrogation but independently send the aircraft's position and velocity twice a second and unique identification every 5 seconds; Figure 1 provides a diagram of the system. It is currently in the roll-out phase, but as of today ADS-B is already operational within the Atlantic airspace and is integrated into modern collision avoidance systems (see Section 4.B.2). It is mandated for use by all aircraft from 2020 in all European and American airspace, and considered a key part of NextGen and SESAR [12].

The rise of SDRs has increased concerns about the security of ADS-B. Modern attacks only require standard COTS hardware to execute, as demonstrated in [8]. Trivially injected ADS-B messages claiming to be non-existing aircraft are impossible to distinguish from authentic ones on the link layer, regardless of the placement of the attacker. To conduct such an attack, it is sufficient to have a line of sight connection from the attacker to the legitimate ADS-B receivers operated by ATC, which are typically located in known positions on the ground at an airport or Area Control Centre.

4 TCAS is part of a larger family of technologies known as Airborne Collision and Avoidance System (ACAS)

Although in many cases redundant systems (such as multilateration) could help mitigate this isolated attack, this is unaccounted for in current standards and left to the implementation of every ADS-B user. Other attacks virtually modify the trajectory of an aircraft by selectively jamming an aircraft's messages and replacing them with modified data. This causes discrepancies between the real position and the one received by ATC. This is a worrying prospect, as ADS-B is set to be the main ATC protocol in the long term, with the FAA considering elimination of Mode A/C/S transponders at some point in the future [13].

FIGURE 1: ADS-B SYSTEM DIAGRAM



ADS-B is an example of a digitally networked surveillance protocol causing a move in the balance of power. Even script kiddies and honest-but-curious threat agents such as the hobbyist can exploit the protocol with commodity hardware able to send and receive on 1090 MHz and a range of open source decoders such as dump1090 [14]. Using the same tools, more capable and aggressive threat agents such as cyber terrorists and cyber criminals could launch attacks with relative ease. Works such as [8] describe in detail how attacks could take place with equipment costs in the thousands of dollars. Although attacks are theoretically cheap on ADS-B, if data fusion with other surveillance systems were used, then attacks would be required on all systems, increasing complexity for the threat agent. This would put it out of the reach of less sophisticated hobbyists and potentially even only in the reach of nation state attackers, depending on the resilience of the most secure technology.

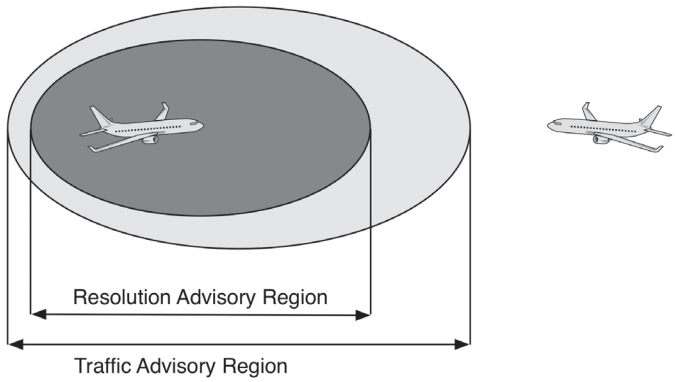
2. TCAS

TCAS allows aircraft to interrogate nearby aircraft in order to resolve airspace conflicts. For example, should another aircraft (referred to as the intruder) come within some predefined

range, TCAS will initially produce a Traffic Advisory (TA) notifying the pilot of traffic nearby. Should the intruder enter the immediate airspace of the aircraft, a Resolution Advisory (RA) will be produced which instructs one of the aircraft to change course. These regions are shown in Figure 2. Usually, the crew will have around 15 seconds to make this change. From 1st December 2015, TCAS is mandated for inclusion on civil aircraft carrying more than 19 passengers or with a minimum take-off weight of 5,700kg [15].

Since TCAS is based on Mode S, it uses an unauthenticated channel. This means that interrogations or responses can be injected as with ADS-B through the use of SDRs. An exemplary vulnerability would be an attacker causing large-scale interference on the 1090 MHz channel without sending on the target frequency, but on the 1030 MHz interrogation frequency instead. Interrogations are currently limited to a maximum of 250 per second [16], but these restrictions are placed on the interrogators, not on the Mode S transponders in aircraft. TCAS would then struggle to operate in a timely fashion given the noise created by answers from surrounding aircraft.

FIGURE 2: TCAS ALERT REGIONS (SIMPLIFIED)



TCAS is also an example of where the interaction of insecure systems produces concerning results. TCAS II, the most modern version, has an optional capability for hybrid surveillance in which ADS-B data from nearby aircraft is used to judge whether intrusions are likely and thus whether a given aircraft should be monitored. The system reduces the number of interrogations required for TCAS without a loss to safety [17]. However, as discussed above, ADS-B faces a number of security challenges that affect the trustworthiness of the data it reports. Thus, attacks on ADS-B could also affect safety systems such as TCAS.

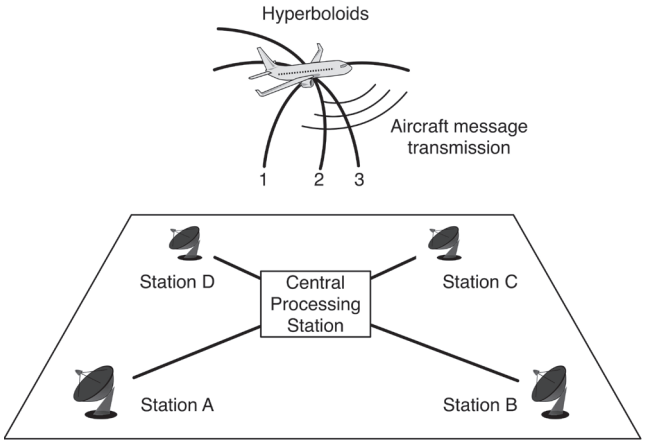
Whilst TCAS is vulnerable due to its design and technologies used, the implications of launching an attack on TCAS are extremely serious. In the best case, ATC may be unable to manage airspace, or aircraft may have a near miss. Jamming the channel or injecting wrong data, however, could cause mid-air collision. As such, we consider that a threat agent who chooses to launch attacks on TCAS would require the motive to cause loss of life or severe disruption, placing it in the domains of nation state and cyber terrorism. However, due to the

lack of control one would have in attacking TCAS, we consider cyber terrorists as the likely threat agent.

3. Multilateration

ADS-B is referred to as 'dependent' due to its dependence on the aircraft in reporting its own measurements such as location and speed. Multilateration provides an alternative method of measuring location and speed without relying on the information reported by the aircraft. Instead, it just relies on receiving a signal from the aircraft, and the Time Difference of Arrival (TDoA) is measured at a number of receivers and a central processing station calculates the transmitter position within a margin of error (see Figure 3).

FIGURE 3: TDoA MULTILATERATION – THE INTERSECTION OF HYPERBOLOIDS 1-3 CALCULATED FROM THE FOUR RECEIVERS A-D REVEALS WHERE THE SIGNAL ORIGINATED



Wide Area Multilateration (WAM) is particularly useful for ATC since it allows location estimation of an aircraft using 1090 MHz messages over large areas. WAM, combined with ADS-B, will form a key part of the next generation surveillance technologies [18] and can help to detect unusual ADS-B reports.

WAM does have a number of challenges of its own, mostly in operation [19]. Due to the number of sensors and data processing equipment required to cover large areas, the cost of installation is very high. As one of the main drivers for ADS-B surveillance is its low cost, this is at odds with WAM.

Due to the inherent redundancy of WAM, attacks would be very costly and resource intensive, which likely makes them possible only for nation states (see Table 3). To spoof an aircraft, one would need to be able to spoof to any receivers in range with perfect timing, and the set of receivers will change as the aircraft moves. This makes WAM very hard to attack, and we consider it to only be in reach of nation states or similarly capable actors.

Table 3 summarises the capabilities of the different threat agents and the surveillance systems that are of interest to them, further estimating the possible cost of the required hardware.

TABLE 3: OVERVIEW OF ATTACKER CAPABILITIES

Threat Agent	Capabilities	Hardware / Cost	Systems of Interest
Passive Observers	Eavesdropping, use of website & mobile apps.	Internet access, $10 SDR receiver stick	ADS-B, Mode S
Script Kiddie / Hobbyist	Eavesdropping, replay attacks, denial of service.	COTS SDR transmitter, $300-$2,000.	ADS-B, Mode S
Cyber Crime	Resources for large-scale operations with sophisticated transponders.	Directional antennas, small UAVs with SDR transmitters, $5,000-$10,000.	ADS-B, Mode S
Cyber Terrorism	Resources for specific high-impact operations, though usually on a limited scale	As with cyber crime but potentially on a smaller, more targeted scale.	ADS-B, TCAS, Mode S
Nation State	Anything physically and computationally possible.	Military-grade radio equipment, capability for electronic warfare.	Any ATC system

5. REASONS FOR THE CURRENT STATE OF WIRELESS SECURITY IN AVIATION

After providing an exemplary overview of wireless security in aviation with our case study, we investigate the underlying reasons of how the current situation came to be. We identify five causes that have led to the apparent lack of communications security within the air traffic system, and which explain the difficulties in fixing it quickly.

A. Long development and certification cycles

The development and certification cycles for new technologies in aviation are typically up to two decades. Taking ADS-B as our running example, the development of its current form started in the late 1990s. The widespread rollout and mandatory use will however only be completed by 2020 in the most advanced airspace. This slow and cautious approach reflects the safety-focused thinking within the aviation community, where a multitude of tests and certifications are required before giving a technology the green light. Unfortunately, while this approach is extremely effective in reducing technical failures, it does not take into account the increased adversarial potential and shifting threat model created by the recent advances in wireless technologies discussed in Section 2.

B. Legacy and compatibility requirements

As a truly global and interconnected system, civil aviation requires technical protocols and procedures that are understood as widely as possible. New protocols and technical advances are not introduced in all airspace at the same time, but depend on local authorities and available infrastructure. It follows that older technologies are kept in service not only for backup reasons, but also to offer the largest compatibility for air traffic control all over the world.

C. Cost pressures

Tying into the previous point, the aviation industry is famously competitive and under major cost pressures [20]. Changes to existing aircraft equipment are expensive and thus unpopular unless they provide immediate cost or operational benefits to the aircraft operators who foot the bill for the installation of new technologies. Apart from these two main drivers, fundamental equipment changes happen primarily through regulatory directives, which are often subject to long lead times and struggle with extensive industry lobbying. As a compromise, legacy technologies are sometimes overhauled to save costs. An example for this is the ADS-B protocol developed in the 1990s, which relies on the old Mode S technology instead of using a new data link (such as the Universal Access Transceiver, or UAT) that was developed from the bottom up.

D. Frequency overuse

As shown in [12] and [21], some of the ATC frequencies such as the 1090 MHz channel are severely congested. An ever-increasing number of aircraft share the same frequencies, exacerbated by UAV set to enter the controlled airspace in the near future. As a consequence, existing ATC protocols suffer from severe message loss, inhibiting potential cryptography-based security solutions at the same time.

E. Preference for open systems

There is a case for air traffic communication protocols to be open to every user; while authentication would be highly desirable, confidentiality through encryption of the content would not. Despite the associated security and privacy problems, the International Civil Aviation Organisation (ICAO) plans for future protocols to be openly accessible. This approach is supposed to fulfil typical aviation requirements such as ease of communication, compatibility, and dealing with administrative differences across countries and airspace [22]. While we acknowledge that open systems are a requirement for the effectiveness of air traffic control for the foreseeable future, it is crucial to start considering and mitigating the downsides, which are rapidly increasing due to previously discussed technological changes.

A further complication for the use of cryptographic means to secure air traffic communication is the inherent complexity of public key infrastructures (PKI). While there are military equivalents to civil SSR in use and under development (STANAG 4193 / Mode 5), due to obvious secrecy requirements, very few details are publicly available. The main problem for a PKI to solve is key distribution and management, [10], which is comparatively easy in closed military environments but very difficult in the open and worldwide system of civil aviation, also tying into the point on compatibility requirements.

A PKI shared by all countries in the world (presumably through ICAO and national agencies) is a monumental task for which no proper suggestions yet exist, and the creation of entirely new protocols is certainly required. The 112 bit message size of ADS-B is too small even for today's computing capabilities, let alone future capabilities; keys would be broken in seconds [10]. While certainly not impossible, experiences with the Internet have also shown that PKI certificate breaches are very common, leaving us with no great solution to the problem.

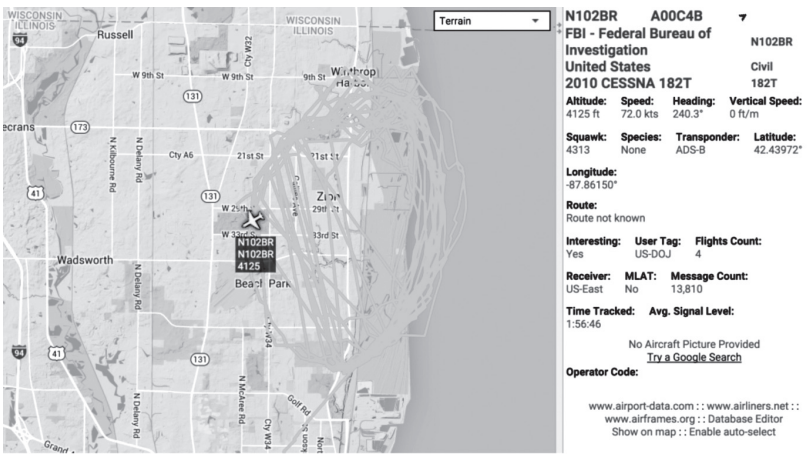
6. AVIATION COMMUNICATIONS SECURITY AS A NEW CYBER POWER ISSUE

Open systems and insecure wireless technologies raise concerns that go beyond active attacks on critical infrastructures: powerful actors increasingly lose their informational edge and privacy as aircraft information becomes available widely and easily on the Internet. We postulate that the democratisation of information has led to a partial erosion of power for state actors, as airborne missions become known immediately to even passive observers.

Traditional 'offline' solutions to maintaining privacy and secrecy for aircraft such as the ASDI scheme [23], which prevents the public display of aircraft movements, have become long obsolete in the SDR era. Plane-spotters around the world detect anomalies, potential incidents, and 'interesting aircraft' practically immediately, and on a large scale, a capability previously limited to state actors. Social media accounts tracking emergency broadcasts provide instant news coverage for both the press and interested individuals, much to the chagrin of some in the traditionally closed aviation community. Hijacked airplanes, too, are detected easily by individuals at home and shared in real-time over Twitter while the aircraft is still in the air [24].

The same effect can be observed for intelligence and security services. With increasing automation and availability of online aviation feeds, the development has gone from occasional sightings of aircraft operated by domestic security services to the large-scale and immediate detection of all transponder-equipped aircraft. An example of the implications of this technological shift is the recent uncovering of a large number of surveillance aircraft employed through front companies of the FBI, an operation that had previously gone unnoticed for some decades [25]. While some of the largest online flight tracking services such as FlightRadar24 comply with requests to not display private or sensitive aircraft data, there are many unregulated sources available that clearly identify such aircraft as interesting to the public (see Figure 4).

FIGURE 4: TRACKING A DOMESTIC SURVEILLANCE AIRCRAFT ON ADSBEXCHANGE.COM

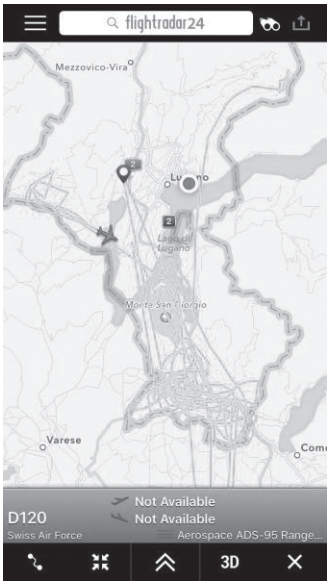


In military settings, this type of open surveillance using data gleaned from Mode S and ADS-B broadcasts has led to similar information leakage through the intentional or unintentional use of transponders during active missions. The diligent tracking of recent airborne engagements in Syria by NATO and Russian aircraft illustrate this point [26]. As airstrikes and reconnaissance missions can easily be detected and anticipated, potentially sensitive strategic and operational information is broadcasted, and deniability of airborne actions becomes difficult; the impact of insecure civil aviation protocols on military users is growing.

The advance of UAVs might offset some or all of this loss of power, as manned missions are replaced with more covert drones. However, in non-military settings, the same problems that are causing concern for current manned surveillance aircraft also apply to UAVs. As aviation authorities are expected to maintain a similar standard of rules for drones in civil airspace, the mandatory use of ADS-B transponders will retain the broadcasts of sensitive data to anyone who is listening.

As an example, Swiss military drones are forced to fly with their ADS-B transponders on when they perform surveillance missions searching for criminals and border breaches [27]. Human traffickers and smugglers can easily track their positions using their smartphones, and avoid discovery by moving only when the drones are not operating.

FIGURE 5: TRACKING BORDER SURVEILLANCE UAV IN REAL TIME USING A MOBILE APP [27]



These examples illustrate the impact that merely passive threat agents have currently. Active attacks on wireless air traffic communication protocols, such as the possibilities discussed in Section 4, could have much greater effects on critical infrastructures in the future.

7. RELATED WORK

Many critical infrastructure industries besides aviation have to adapt to a shifting threat model caused by the rapid advance of technology. We briefly discuss some of the work related to ours in this section.

In the area of transport infrastructure, recent work has shown that current cars use weak authentication and often offer no integrity mechanisms for their on-board systems. Koscher et al. [28] demonstrated this on car data networks even as attempts at using security were being made. This is a dramatic shift whereby cars are now attackable via computer systems, with which the automobile industry has not yet dealt. When scaled up to public transport such as trains, we see the inclusion of industrial control systems (ICS). As demonstrated by Igure et al. [29], the rise of conventional networking technology in ICS without proper security has led to a range of new challenges. Typically, these are similar to those faced in aviation and automotive, such as a lack of authentication or integrity of data networks. Unlike aviation, however, the Repository of Industrial Security Incidents database [30] indicates that attacks on these systems are already occurring. This indicates that, given the opportunity, attackers will exploit these vulnerabilities.

As the use of COTS technologies increases in aviation, scenarios such as those seen in ICS become more common. This has led to a number of works addressing aviation security at a conceptual level. For example, McParland [31] discusses how cryptography can help in protecting the Aeronautical Telecommunications Network (ATN) and some of its applications. Stephens [32] provides a range of security methods and primitives used in typical networking scenarios that could be relevant to aviation. More recently, Li et al. [33] propose a security architecture for next generation communications in air traffic control (ATC). It presents a defence-in-depth approach, and extends to navigation and surveillance at a conceptual level, but does not deal with specific systems.

Within the wireless security community, much work has been done on ADS-B, as it provides a popular example of changing threat models rendering next generation systems insecure. Schäfer et al. [8] experimentally test wireless attacks on ADS-B integrity and availability, analysing the power, timing and range constraints in the real world. Strohmeier et al. [12] assess ADS-B as a system including the intended use, current deployment status, and analysis of the channel characteristics. It also analyses the security issues such as ghost aircraft injection at a high level and comprehensively discusses potential security measures for the future.

To the best of our knowledge, ours is the first work to introduce a threat agent model for modern aviation, and to analyse the impact on the cyber power of nation state actors by novel wireless security threats to air traffic communication.

8. DISCUSSION AND CONCLUSION

In this article, we outlined how technological advances change security threat models in aviation and influence the current cyber power balance. We developed a taxonomy of different threat agents able to affect the cyber-physical aviation system. We postulate that the evolution of cyber power of these agents in the present and in the expected future is an important aspect of future cyber conflicts. Advances in wireless technology and increased digitalisation and automation in aviation enable simpler attacks with few resources. This trend moves power away from nation states towards cyber criminals and terrorists, and even unorganised hobbyists or passive observers.

If nation state actors want to restore the previous balance of power, and increase the security of the aviation system, awareness of cyber security issues among aviation circles and governments is a key factor. Only by raising awareness can the necessary research and development happen, enabling the responsible bodies to address the problem, and prevent the exploitation of existing vulnerabilities in the future.

Considering the decades-long development and certification cycles, research on protocols that include security by design is required as quickly as possible even though it will only pay in the long-term. Existing examples of security designs and analyses for the ADS-B protocol (see, e.g., [34]) can inform the directions of such future research.

New protocols can also provide improvements for the issues of aviation privacy and secrecy. With proper design and implementation of pseudonymous identifiers, most of the relevant information leakage could be reduced to the level of previous, non-technologically enhanced, plane-spotting days, particularly concerning military, government, and private aviation.

Finally, we argue that top-down regulations are crucial in an industry such as aviation that is very cost-conscious and where actions are often taken only when required by regulators. Tying in with the previous point about awareness, the authorities need to be put in a knowledgeable position to issue the necessary regulations, and they should further consider the effect of their actions – or inaction – on the future balance of cyber power.

REFERENCES

[1] M. Strohmeier, M. Schafer, V. Lenders, and I. Martinovic, "Realities and challenges of nextgen air traffic management: the case of ADS-B," *IEEE Commun. Mag*., vol. 52, no. 5, pp. 111–118, 2014.

[2] Department of the Army, "FM 3-38: Cyber Electromagnetic Activities," *F. Man*., no. 1, p. 96, 2014.

[3] A. Costin and A. Francillon, "Ghost is in the Air(traffic): On insecurity of ADS-B protocol and practical attacks on ADS-B devices," in *Black Hat USA*, 2012.

[4] "GNU Radio," 2016. [Online]. Available: https://gnuradio.org. [Accessed: 12-Apr-2016].

[5] D. Adamy, *Introduction to electronic warfare modelling and simulation*. SciTech Publishing, 2006.

[6] N. Foster, "gr-air-modes GitHub Repository," 2015. [Online]. Available: https://github.com/bistromath/gr-air-modes. [Accessed: 12-Apr-2016].

[7] J. Croft, "Forensic Mining With ADS-B," *Aviation Week & Space Technology*, 2015. [Online]. Available: http://aviationweek.com/commercial-aviation/forensic-mining-ads-b. [Accessed: 12-Apr-2016].

[8] M. Schäfer, V. Lenders, and I. Martinovic, “Experimental analysis of attacks on next generation air traffic communication,” in *Applied Cryptography and Network Security*, pp. 253–271, 2013.
[9] K. Stouffer, J. Falco, and K. Scarfone, “Guide to Industrial Control Systems (ICS) Security,” *Recomm. Natl. Inst. Stand. Technol*., no. SP 800–82, pp. 1–157, 2007.
[10] M. Strohmeier, V. Lenders, and I. Martinovic, “On the Security of the Automatic Dependent Surveillance-Broadcast Protocol,” *IEEE Surv. Tutorials*, vol. 17, no. 2, pp. 1066 – 1087, 2015.
[11] M. Strohmeier, M. Schäfer, R. Pinheiro, V. Lenders, and I. Martinovic, “On Perception and Reality in Wireless Air Traffic Communications Security.” 2016.
[12] M. Strohmeier, M. Schäfer, V. Lenders, and I. Martinovic, “Realities and challenges of NextGen air traffic management: the case of ADS-B,” *IEEE Commun. Mag*., vol. 52, no. 5, 2014.
[13] Federal Aviation Administration, “ADS-B Frequently Asked Questions,” 2015. [Online]. Available: https://www.faa.gov/nextgen/programs/adsb/faq/. [Accessed: 12-Apr-2016].
[14] “dump1090 GitHub repository,” 2016. [Online]. Available: https://github.com/antirez/dump1090. [Accessed: 26-Dec-2015].
[15] The European Commission, *Commission regulation laying down common airspace usage requirements and operating procedures for airborne collision avoidance*, no. 1332. European Union, 2011, pp. 2008–2010.
[16] Aeronautical Surveillance Panel, “Draft Doc9924 guidance material for the measurement of all-call reply rates,” International Civil Aviation Organisation, 2013.
[17] S. Henely, “Traffic Alert and Collision Avoidance System II (TCAS II),” in *Digital Avionics Handbook*, 3rd ed., C. R. Spitzer, U. Ferrell, and T. Ferrell, Eds. CRC Press, pp. 1–9, 2015.
[18] International Civil Aviation Organisation, “Initial capability for ground surveillance,” in *Global Air Navigation Plan* 2013-20282, 2013, p. 56.
[19] G. Galati, M. Leonardi, P. Magarò, and V. Paciucci, “Wide area surveillance using SSR Mode S multilateration: advantages and limitations,” *Eur. Radar Conf*., pp. 225–229, 2005.
[20] M. Franke, “Competition between network carriers and low-cost carriers—retreat battle or breakthrough to a new level of efficiency?,” *J. Air Transp. Manag*., vol. 10, no. 1, pp. 15–21, 2004.
[21] Eurocontrol, “Updated work on 5 - Final report on electromagnetic environmental effects of, and on, ACAS,” Aug. 2009.
[22] International Civil Aviation Organisation, “Review report of the thirteenth meeting of Automatic Dependent Surveillance-Broadcast (ADS-B) study and implementation task force.” Beijing, 2014.
[23] National Business Aviation Administration, “Blocking display of Aircraft Situation Display to Industry (ASDI) data,” 2016. [Online]. Available: https://www.nbaa.org/ops/security/asdi/. [Accessed: 22-Feb-2016].
[24] J. Walton, “How I broke the Ethiopian Airlines #ET702 hijacking on Twitter,” 2014. [Online]. Available: https://medium.com/@thatjohn/how-i-broke-the-ethiopian-airlines-et702-hijacking-on-twitter-6c2cc1d2f2e4#.pmobgbtqu. [Accessed: 12-Apr-2016].
[25] C. Friedersdorf, “Congress Didn’t Notice the FBI Creating a ‘Small Air Force’ for Surveillance,” *The Atlantic*, 2015. [Online]. Available: http://www.theatlantic.com/politics/archive/2015/06/congress-didnt-notice-the-fbi-creating-a-small-air-force-for-surveillance/395147/. [Accessed: 12-Apr-2016].
[26] D. Cenciotti, “Online flight tracking provides interesting details about Russian air bridge to Syria,” *The Aviationist*, 2015. [Online]. Available: http://theaviationist.com/2015/09/11/ads-b-exposes-russian-air-bridge-to-syria/. [Accessed: 12-Apr-2016].
[27] “App zeigt Kontroll-Flug von Armeedrohne,” *20 Minuten*, 2015. [Online]. Available: http://www.20min.ch/schweiz/news/story/App-zeigt-Kontroll-Flug-von-Armeedrohne-27294424. [Accessed: 12-Apr-2016].
[28] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, and S. Savage, “Experimental security analysis of a modern automobile,” 2*010 IEEE Symp. Secur. Priv*., pp. 447–462, 2010.
[29] V. M. Igure, S. A. Laughter, and R. D. Williams, “Security issues in SCADA networks,” *Comput. Secur*., vol. 25, no. 7, pp. 498–506, 2006.
[30] Exida LLC, “Repository of Industrial Security Incidents (RISI) Online Incident Database,” 2015. [Online]. Available: http://www.risidata.com/Database. [Accessed: 12-Apr-2016].
[31] T. McParland and V. Patel, “Securing air-ground communications,” in *20th Digital Avionics Systems Conference,* 2001, pp. 1–9.
[32] B. Stephens, “Security architecture for aeronautical networks,” in *23rd Digital Avionics Systems Conference*, 2004, vol. 2.
[33] W. (Wenhua) Li and P. Kamal, “Integrated aviation security for defense-in-depth of next generation air transportation system,” *2011 IEEE Int. Conf. Technol. Homel. Secur*., pp. 136–142, 2011.

[34] C. Finke, J. Butts, R. Mills, and M. Grimaila, “Enhancing the security of aircraft surveillance in the next generation air traffic control system,” *Int. J. Crit. Infrastruct. Prot.*, vol. 6, no. 1, pp. 3–11, Mar. 2013.
[35] J. Carney, “Is Spying on Corporate Jets Insider Trading?,” *CNBC*, 2012. [Online]. Available: http://www.cnbc.com/id/100272132. [Accessed: 12-Feb-2016].

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Countering Advanced Persistent Threats through Security Intelligence and Big Data Analytics

Mirco Marchetti
Department of Engineering 'Enzo Ferrari'
University of Modena and Reggio Emilia
Modena, Italy
mirco.marchetti@unimore.it

Fabio Pierazzi
Department of Engineering 'Enzo Ferrari'
University of Modena and Reggio Emilia
Modena, Italy
fabio.pierazzi@unimore.it

Alessandro Guido
Department of Engineering 'Enzo Ferrari'
University of Modena and Reggio Emilia
Modena, Italy
alessandro.guido@unimore.it

Michele Colajanni
Department of Engineering 'Enzo Ferrari'
University of Modena and Reggio Emilia
Modena, Italy
michele.colajanni@unimore.it

Abstract: Advanced Persistent Threats (APTs) represent the most challenging threats to the security and safety of the cyber landscape. APTs are human-driven attacks backed by complex strategies that combine multidisciplinary skills in information technology, intelligence, and psychology. Defending large organisations with tens of thousands of hosts requires similar multi-factor approaches. We propose a novel framework that combines different techniques based on big data analytics and security intelligence to support human analysts in prioritising the hosts that are most likely to be compromised. We show that the collection and integration of internal and external indicators represents a step forward with respect to the state of the art in the field of early detection and mitigation of APT activities.

Keywords: *Advanced Persistent Threats, big data analytics, security intelligence, social engineering*

1. INTRODUCTION

The majority of cyber-attacks rely on automated scanning and exploitation of known vulnerabilities over large sets of targets. APTs represent a more dangerous category because they are sophisticated human-driven attacks against specific targets. Objectives of APT attacks include continuous exfiltration of information, cyber warfare, damage to critical infrastructure, and degradation of military assets (Data Breaches, 2016). They are typically perpetrated over long periods of time by groups of experts that leverage open source intelligence and social engineering techniques (Molok, Chang, & Ahmad, 2010), vulnerabilities not always known to the public (Jeun, Lee, & Won, 2012; Virvilis, Serrano, & Vanautgaerden, 2014), standard protocols, encrypted communications, and zero-day vulnerabilities to evade detection (Brewer, 2014). Consequently, traditional defensive solutions such as antiviruses and signature-based detection systems (Sabahi & Movaghar, 2008; Zhou, Leckie, & Karunasekera, 2010) that can identify standard malware are ineffective against APTs.

We claim that effective defences require analogous multi-factor approaches where human analysts must be supported by big data analytic techniques that are able to detect and prioritise weak signals related to APT activities. The proposed AUSPEX[1] framework follows these principles. It gathers and combines *internal information* from network probes located in an organisation, and *external information* from public sources such as the web, social networks, and blacklists. From these data, AUSPEX calculates two sets of indicators:

1. *compromise indicators*, that prioritise internal clients based on their suspicious network activities; and
2. *exposure indicators*, that estimate the likelihood of a social engineering or intelligence attack.

The final output of AUSPEX is a list of internal hosts ranked by *compromise* and *exposure* scores. In this version, AUSPEX focuses on client hosts that are likely the initial targets of APTs.

Papers related to APTs usually focus on the most popular attacks (Brewer, 2014; Jeun, Lee, & Won, 2012; Virvilis & Gritzalis, 2013) and identify the main phases of an APT without proposing detection approaches. Other works (Bhatt, Toshiro Yano, & Gustavsson, 2014; Giura & Wang, 2012; De Vries, Hoogstraaten, van den Berg, & Daskapan, 2012; Hutchins, Cloppert, & Amin, 2011) formalise the APT defence problem, but they only propose development guidelines and leave the definition of detection rules and analysis to future work. To the best of our knowledge, AUSPEX is the first framework that supports human analysts to detect and mitigate APTs in large organisations by prioritising weak signals through a combination of internal and external data.

1 An *auspex* was an interpreter of omens in ancient Rome.

The remainder of the paper is organised as follows. Section 2 introduces the scenario and main challenges related to APT identification. Section 3 presents an overview of the AUSPEX framework. Sections 4 and 5 discuss *compromise* and *exposure* indicators, respectively. Section 6 shows how hosts are ranked, section 7 compares AUSPEX with related literature, and section 8 draws conclusions and outlines future work.

2. ADVANCED PERSISTENT THREATS

A typical APT attack comprises five main phases (Brewer, 2014): reconnaissance; compromise; maintaining access; lateral movement; and data exfiltration.

In the *reconnaissance* phase, an attacker carries out intelligence analysis on the target organisation to extract information and identify weak spots (Lindamood, Heatherly, Kantarcioglu, & Thuraisingham, 2009; Irani, Webb, Pu, & Li, 2011). This phase involves both social and technological aspects.

In the *compromise* phase, an APT attacker infiltrates the system, possibly through social engineering strategies that exploit information gathered during the reconnaissance phase. Often, this phase involves infected files sent as email attachments or links (Data Breaches, 2016). The final goal is to install a RAT (Remote Access Trojan, or Remote Administration Tool) on at least one host of the organisation.

In the *maintaining access* phase, an attacker uses the RAT to communicate with an external *Command and Control* (CnC) server (Bailey, Cooke, Jahanian, Xu, & Karir, 2009). An internal host of the organisation initiates this communication, because outgoing traffic passes more easily through firewalls.

The *lateral movement* phase has two main purposes: to shift towards other internal hosts of the organisation that have more access privileges or intrinsic value; and to move data to a drop zone, such as a web server, that allows information exfiltration while minimising risks of detection.

Finally, in the *data exfiltration* phase, the attacker uploads data to an external server, either in a single burst or slowly over several days. The attacker may also use legitimate external servers as drop zones, such as cloud hosts.

The challenge of APT detection is inherent to the combined use of different attack vectors and evasion strategies. Therefore, defences cannot be purely technological and must be based on a combination of human analysis and automatic detection of weak signals likely characterising an APT. An overview of the proposed defensive approach is described in the following section.

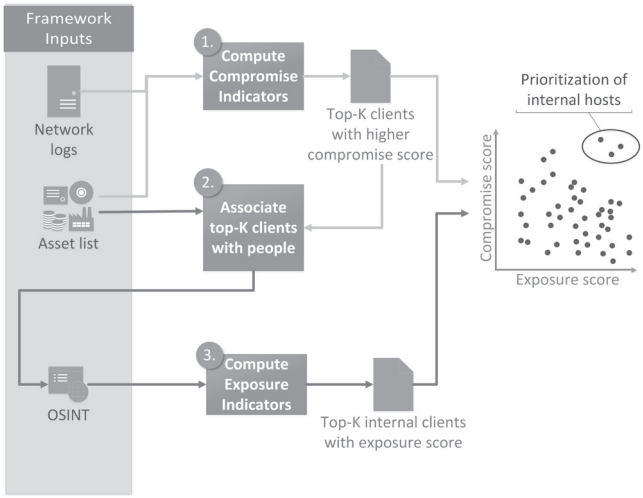
3. OVERVIEW OF THE DEFENSIVE METHOD

The main purpose of AUSPEX is to prioritise the internal clients of the organisation that are most likely compromised by an APT. To this end, it combines weak signals derived by security sensors with external information related to employees who may be victims of social engineering attacks. Figure 1 shows an overview of the main AUSPEX components.

The input data gathered and analysed by AUSPEX are shown on the left column of Figure 1:

- **Network logs** (e.g., network flows, requests to web servers, requests to name servers, security alerts) collected through SIEM and intrusion detection systems (Denning, 1987; Paxson, 1999). This data allows the identification of weak signals possibly corresponding to APT activities.
- A simplified **assets list** that maps the members of the organisation and their client devices. This information is useful to *link* technological and open source data.
- **OSINT information** collected from public open sources. This is used to identify and quantify the information that is available to APT attackers.

FIGURE 1: AUSPEX OVERVIEW



AUSPEX adopts a network-centric approach because network traffic can be collected and analysed more easily than host-based logs (Friedberg, Skopik, Settanni, & Fiedler, 2015) (e.g., OS system calls), especially in large and dynamic organisations comprising heterogeneous hardware and software components. Hence, as a first step (box 1 in Figure 1) AUSPEX analyses network logs collected within the organisation to evaluate a set of compromise indicators for each internal client. For example, it is possible to estimate the probability of data exfiltration by analysing outgoing traffic of each internal host. Then, an overall *compromise score* is calculated from the set of indicators. This part is described in Section 4. The clients with higher

compromise scores (top-K clients) are selected for further analysis on external sources. To this end, AUSPEX uses the list of the top-K clients and the asset list to identify users of the most likely compromised clients (box 2 in Figure 1). For each of these users AUSPEX calculates a set of *exposure indicators* (box 3 in Figure 1) by crawling open sources to understand whether these users might be likely victims of social engineering attacks (Molok, Chang, & Ahmad, 2010). This part is described in section 5.

Computation of the exposure indexes is only performed for the top-K clients with higher compromise scores to reduce the volume of data. Modern organisations usually have thousands of internal devices; hence, there is a need to narrow the scope of an analysis that can be repeated daily. Since the amount of open source data is theoretically unrestrained, focusing attention only on the top-K clients also makes crawling open sources feasible and more effective. Moreover, generalised crawling and inspection of employee information is likely to raise issues related to labour and privacy laws in many countries. Focused analyses on employees using likely compromised hosts will reduce legal risks.

The final output of AUSPEX is a list of internal clients and overall compromise and exposure scores. By prioritising hosts in which both scores are high, the security analysts can focus attention on a limited subset of internal clients.

4. COMPROMISE INDICATORS

Despite the huge challenge of detecting human-driven targeted APTs in large networks, it is possible to define automatic analyses to support security analysts in prioritising events possibly related to APTs (Brewer, 2014; Virvilis & Gritzalis, 2013). To this end, AUSPEX adopts algorithms targeted to prioritise internal clients possibly involved in *maintaining access, lateral movement*, and *data exfiltration* phases of an APT.

4.1 Maintaining access

After initial compromise, an APT attacker deploys a RAT (Remote Administration Tool) on one or more hosts of the organisation. This RAT tries to contact one or more external CnC servers controlled by the APT attacker, hence the first goal is to detect communications towards these CnC servers. To achieve this, AUSPEX analyses communications between internal and external hosts to identify suspicious external hosts and evaluate how many flows exist between them and internal clients.

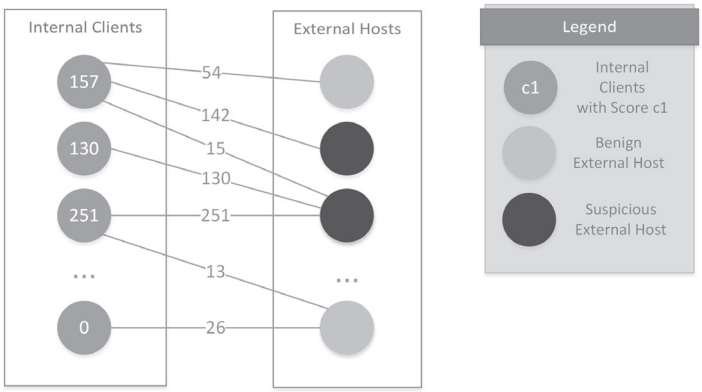
AUSPEX analyses network logs and builds an undirected bipartite graph where internal and external hosts are represented as two sets of vertices. Edges between vertices have weights that correspond to the number of flows between internal and external hosts. It then labels each external host as benign or suspicious, relying on a combination of algorithms including blacklist filtering, DGA analysis, regular access patterns, and non-matching flows (Bailey, Cooke, Jahanian, Xu, & Karir, 2009; Bilge, Balzarotti, Robertson, Kirda, & Kruegel, 2012; Schiavoni, Maggi, Cavallaro, & Zanero, 2014).

Finally, AUSPEX calculates a *CnC* compromise indicator c_1^h for each internal client *h* that corresponds to the *sum of weights* of the edges that connect *h* with external hosts. The subscript 1 indicates that this is the first of three compromise indicators. Other two indicators related to the *lateral movement* and *data exfiltration* phases are presented in sections 4.2 and 4.3. The score c_1^h estimates the likelihood that an internal host *h* is involved in CnC communications.

A simplified example of the proposed algorithm is shown in Figure 2, where internal clients (on the left) communicate with one or more external hosts (on the right). We observe that the internal client at the bottom has score c_1^h=0 because it is communicating with a benign external host.

AUSPEX marks an external host as suspicious if it satisfies at least one of the four criteria (Bilge, Balzarotti, Robertson, Kirda, & Kruegel, 2012; Bailey, Cooke, Jahanian, Xu, & Karir, 2009), *blacklist filtering, DGA analysis, regular access patterns, and non-matching flows.*

FIGURE 2: EXAMPLE FOR THE ALGORITHM FOR CNC COMMUNICATION RANKING



Blacklist filtering. Several public sites offer reputation scores for IP addresses and domain names that represent their likelihood of being malicious. Examples are *Malware Domain List*,[2] *WhatIsMyIP Blacklist Check*,[3] *Google Safe Browsing*,[4] and *Web of Trust*.[5] AUSPEX calculates a reputation score through the equation proposed in (Bilge, Balzarotti, Robertson, Kirda, & Kruegel, 2012).

DGA analysis. Domain-Generation Algorithms are often adopted by attackers to simplify communications between RATs and CnCs (Bailey, Cooke, Jahanian, Xu, & Karir, 2009). To determine whether a domain name is DGA-generated, AUSPEX adopts a simplified version of the algorithm proposed in (Schiavoni, Maggi, Cavallaro, & Zanero, 2014). The main intuition is

2 Malware Domain List homepage, [Online]. Available: http://www.malwaredomainlist.com, [Accessed 13.04.2016].
3 WhatIsMyIP.com's Blacklist Check, [Online]. Available: https://www.whatismyip.com/blacklist-check, [Accessed 13.04.2016].
4 Safe Browsing in Google's Transparency Report, [Online]. Available: https://www.google.com/transparencyreport/safebrowsing, [Accessed 13.04.2016].
5 Web of Trust homepage, [Online]. Available: https://www.mywot.com, [Accessed 13.04.2016].

that a domain name is likely benign (not a DGA) if it is composed of meaningful words (those in the English lexicon) and is pronounceable. Hence, AUSPEX calculates two main metrics for each domain *d:* the *meaningful characters ratio R(d)* and the *n-grams normality score* $S_n(d)$.

Then, a feature vector $f(d)=[R,S_1,S_2,S_3]$ is associated with each domain *d*. The training of benign domains is done on the Alexa top 100,000 sites list, where a centroid $\bar{x}$ of the feature space is determined. Then, a domain is marked as DGA if the *Mahalanobis distance* (Bishop, 2006) between a feature *f(d)* and the centroid $\bar{x}$ of the benign feature space is higher than a threshold, Λ, defined as the *p*- percentile of the distance vectors values from the centroid. Based on our experience and the literature (Schiavoni, Maggi, Cavallaro, & Zanero, 2014), we suggest setting Λ=0,9.

Regular access patterns. If access patterns between an external host and one or more internal hosts are too regular, they likely correspond to automatic communications. As proposed in (Bilge, Balzarotti, Robertson, Kirda, & Kruegel, 2012), AUSPEX calculates a *set of inter-arrival sequences* $I(h_{ext})$ as the union of the differences between the timestamps of contiguous flows between h_{ext} and each internal client:

$$I(h_{ext}) = \bigcup_{int} \left(\bigcup_{k=1}^{K_{int,ext}} (ts_{int,ext,k} - ts_{int,ext,k-1}) \right)$$

where $K_{int,ext}$ is the number of flows between the external host h_{ext} and internal host h_{int}, and $ts_{int,ext,k}$ is the value of the timestamp corresponding to flow number *k* identified between h_{ext} and h_{int} . An external host is marked as suspicious if there is an inter-arrival difference value $\bar{n}$ with probability higher than a threshold $\bar{P}$ (meaning that too many communications occur at regular intervals of duration $\bar{n}$). Based on our test deployments in real and large network environments, we recommend initially setting $\bar{P}$=0,95, that is, a probability of 0.95.

Non-matching flows. If there is an imbalanced number of flows between an external host and one or more internal hosts h^1_{int} , h^2_{int} , ... , h^N_{int} , it is possible that an external CnC is no longer reachable, and one or more internal clients are still trying to contact it. To quantify this, AUSPEX adapts a metric initially proposed in (Bilge, Balzarotti, Robertson, Kirda, & Kruegel, 2012). An external host h_{ext} is marked as suspicious if it has responded to less than 50% of the flows.

4.2 Lateral movement

In the *lateral movement* phase, an APT attacker that has infiltrated an organisation's network tries to gain access to other internal hosts (clients or servers) to improve his chances of persistence and to acquire greater privileges for accessing resources of interest.

To identify revealing signals of ongoing lateral movements, AUSPEX analyses internal communications, defined as any network activity or interaction occurring between two internal

hosts. Internal hosts of an organisation belong to two groups: clients assigned to employees, and servers (e.g., a Network Attached Storage or an HTTP server). Since AUSPEX focuses on prioritising clients, we only analyse client-to-server and client-to-client communications.

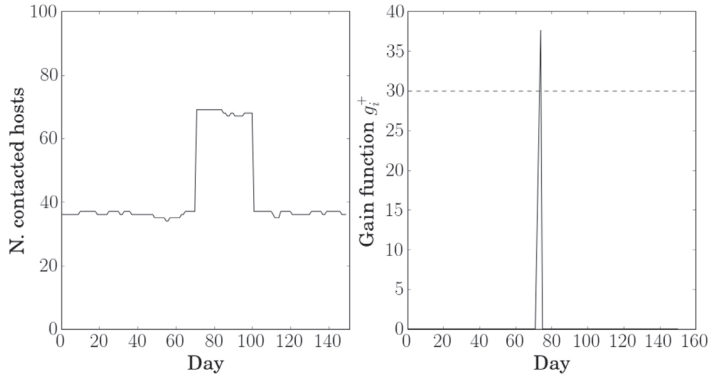
We analysed the network traffic generated within the internal network of a real-world organisation. Our analyses confirmed that the number of packets exchanged between internal hosts has a high variance, and greatly depends on human activities. We also observed that for each internal client h, the number of internal hosts contacted by is stable, because a client tends to communicate with a stable set of internal hosts. Hence, to identify possible lateral movements, for each internal client h AUSPEX monitors the number of internal hosts (both clients and servers) that have communicated with h in the recent past. Significant changes in this value imply that h is contacting many new internal hosts, an activity that is likely related to lateral movements. To this purpose, AUSPEX calculates the time series of the number of internal hosts contacted by h in a sliding time window Δ. This time series is denoted by D_t^h. By monitoring state-changes (Montgomery, 1991) in D_t^h, AUSPEX is able to detect internal clients with sudden increases in the number of contacted hosts.

For example, let us consider an internal client h at day t, that in the previous window Δ=15 days communicated with the set of 5 hosts $\{h_1, h_2, h_3, h_4, h_5\}$. Let us suppose that t at day t+1 contacted the set of hosts $\{h_4, h_5, h_6, h_7, h_8\}$. Although the internal hosts are still five in total, there are three new internal hosts (h_6, h_7, h_8) that h did not contact in the previous window. Hence $D_{t+1}^h=5+3=8$, and a state-change can be detected since the number of hosts contacted by h has almost doubled in one day. AUSPEX adopts a CUSUM-based state-change detection algorithm that monitors the *mean* of the series of contacted hosts (Montgomery, 1991). This family of state-change detectors is applicable to series with low variability and is computationally feasible even for online detection contexts. For each internal host, AUSPEX calculates a *lateral movement* indicator c_2^h as:

$$c_2^h \text{ as: } = \Sigma_t \, g_t^+$$

where g_t^+ is a gain function higher than 0 if a positive state-change is detected at time t, and estimates the entity of the state-change (Casolari, Tosi, & Lo Presti, 2012). An example of state-change observed in a segment of an internal network environment is proposed in Figure 3, where Figure 3a shows an example of the time series D_t^h, and Figure 3b shows the magnitude of the state-change when it is detected around day 70, with Δ=30 days. High values of c_2^h imply that internal client h has frequently or significantly increased the number of internal hosts that it has contacted.

FIGURE 3: EXAMPLE OF STATE-CHANGE DETECTION THROUGH THE CUSUM-BASED ALGORITHM



4.3 Data exfiltration

As a final step of an APT campaign oriented to data exfiltration, the attacker must send confidential data from the target organisation to one or more remote servers. Some recent and popular examples of data exfiltration are: the *Adobe* leak in 2013, comprising 9GB of encrypted password; the *Ashley Madison* leak in 2015, in which their entire database of about 30GB was stolen (Data Breaches, 2016); and the *Hacking Team* data leak, including 400GB of corporate data.

To identify internal hosts possibly involved in data exfiltration, AUSPEX focuses on the analysis on outgoing traffic (Brewer, 2014). For each internal client *h*, AUSPEX calculates a feature vector $x_t=(x_t^1, x_t^2, x_t^3)$, with the following three components: *outbytes* (x_t^1), which captures deviations in the number of bytes uploaded by to external hosts; and *numdst* (x_t^2), and *numconns* (x_t^3), which identify variations in the number of destinations and connections to external hosts. The choice of the right time granularity *t* for the feature vectors x_t is context-dependent (Brockwell & Davis, 2013; Pierazzi, Casolari, Colajanni, & Marchetti, 2016). In general, we recommend a time granularity of 1 day because daily aggregation reduces noise related to normal use of clients and servers. It also allows security analysts to investigate suspicious activities on a daily basis.

AUSPEX then quantifies the suspiciousness of outgoing traffic for each internal host with respect to other internal hosts and their past behaviour. A commonly adopted method to estimate the variation of behaviour is to consider the movement vector as a Euclidean difference in the feature space (Bishop, 2006). To consider the past behaviour of an internal host, AUSPEX also calculates the centroid of its past positions in a time window *W* as follows:

$$\beta_t(W) = \left(\frac{\sum_j x_j^1}{W}, \frac{\sum_j x_j^2}{W}, \frac{\sum_j x_j^3}{W}\right), j \in \{t-W-1, \ldots, t-1\}$$

where the three components of $\beta_t\ (W)$ correspond to the mean of the last W values of the three components of the feature vector x_t. This metric represents an average of the history of uploaded bytes, number of connections, and number of destinations contacted by an internal client. Finally, for each internal client AUSPEX calculates a compromise indicator c_3^h as the magnitude of the movement vector m_t (Bishop, 2006):

$$c_3^h = ||m_t|| = \sqrt{\sum_{i=1}^{N} \left(\frac{x_t^i - \beta_t^i(W)}{\beta_t^i(W)} \right)^2}$$

where m_t quantifies changes in the outgoing traffic statistics. High values of c_3^h imply suspicious uploads that differ significantly with respect to past behaviour.

5. EXPOSURE INDICATORS

Clients with higher compromise scores (top-K clients) are selected for further analysis based on external sources. The goal is to verify whether employees that are likely victims of social engineering attacks use these clients. To this purpose, AUSPEX calculates a set of *exposure indicators* for the employees that use likely-compromised clients by analysing information collected from public and open sources that are also available to an APT attacker. In the present version, AUSPEX considers three popular social networks (Facebook, Twitter, and LinkedIn) that allow fee-based APIs to social profiles and pages. Twitter exposes REST APIs[6] to search for users by name, and to gather information about users, their tweets and their followers. LinkedIn offers APIs to search for people by name[7] or company,[8] and access all metadata, profile information, and public posts of a person. Facebook uses a proprietary query language through the search box of its website. To leverage this interface, AUSPEX adopts Selenium,[9] a software library that handles browser interactions programmatically. In particular, it gets Facebook user IDs through their email address by means of the search bar, and then crawls public data related to user profiles, such as posts and friends lists.

For the employees of the organisation that are using likely compromised clients, AUSPEX calculates a set of exposure indicators that determine the likelihood of a social engineering attack: social activity; social connections; personal information leakage; and organisation information leakage.

The **social activity** indicator e_1 quantifies how much an employee is active on social networks (Romero, Galuba, Asur, & Huberman, 2011; Montangero & Furini, 2015; Canali, Casolari, & Lancellotti, 2010). Employees with higher social activity are more likely to become victims of social engineering attacks, because they may be approached by an APT attacker or may reveal

6 Twitter REST APIs, [Online]. Available: https://dev.twitter.com/rest/public. [Accessed 13.04.2016].
7 LinkedIn Profile API for developers, [Online]. Available: https://developer-programs.linkedin.com/documents/profile-api [Access to this resource requires a valid LinkedIn account.].
8 LinkedIn People Search API for developers, [Online]. Available: https://developer-programs.linkedin.com/documents/people-search-api. [Access to this resource requires a valid LinkedIn account.].
9 SeleniumHQ homepage, [Online]. Available: http://www.seleniumhq.org/. [Accessed 13.04.2016].

sensitive information. AUSPEX evaluates this indicator by considering the average number of posts (i.e., any form of user-generated content posted on the social network) per day on Facebook, Twitter, and LinkedIn.

The **social connections** indicator e_2 quantifies how much an employee is connected to other employees of the organisation on social networks. This information is useful to an attacker because employees connected to many other members of the organisation are likely to know information related to confidential projects. It might also be easier for an APT attacker to propagate through lateral movement to other hosts, such as through infected email attachments. AUSPEX performs this by determining for each employee the number of connections on Facebook, LinkedIn, and Twitter who work for the same organisation.

The **personal information leakage** indicator e_3 corresponds to the number of personal fields that the employee filled in his online social network profiles. The main motivation is that an employee might reveal personal information that attackers may leverage to carry out a more effective social engineering activity. The problem of personal information leakage has already been studied in the literature (Molok, Chang, & Ahmad, 2010; Irani, Webb, Pu, & Li, 2011; Lindamood, Heatherly, Kantarcioglu, & Thuraisingham, 2009; Lam, Chen, & Chen, 2008). AUSPEX considers the following fields of Twitter, Facebook and LinkedIn profiles: name, location, sex, relationship status, hometown, homepage, birthdate, 'about me', groups, interests, liked pages.

The **organisation information leakage** indicator e_4 quantifies the amount of information about the organisation that an employee has published on open sources, and that is publicly available. This is relevant because an APT attacker may select his target based on such information (e.g., target an employee that is working on a specific project). AUSPEX adopts a variation of the score defined in (Irani, Webb, Pu, & Li, 2011) that considers a set of *keywords* that refer to activities of the organisation, such as projects, customers, or suppliers. Each keyword has a risk level, representing its severity. Risk levels can be assigned according to risk assessment best practices (Ostrom & Wilhelmsen, 2012). Examples of risk levels related to a project name are proposed in Table 1.

TABLE 1: EXAMPLE OF RISK LEVELS RELATED TO KEYWORDS

Risk level	Explanation
1 [low]	Public mentions of the project do not affect the organisation. However, the knowledge that an individual is working on this project may increase the chances for an APT attacker to target him.
2 [medium]	The project is known by the members of the organisation, but is not publicly disclosed on the outside.
3 [high]	This project is known only to some members of the organisation. Its diffusion may have moderate legal and economic consequences.
4 [critical]	This project is extremely critical and confidential. Its diffusion may have severe legal and economic consequences.

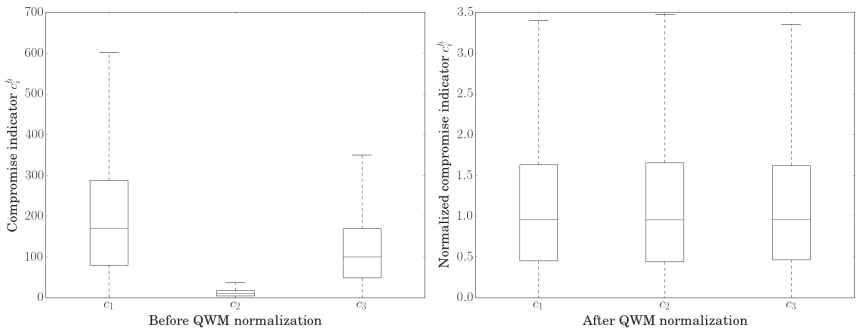
Finally, the organisation information leakage indicator e_4 associated with an employee is determined as the sum of the risk levels associated with keywords that they use. To associate all the exposure indicators to an internal client, AUSPEX uses the *internal assets* list described in Section 3.

6. PRIORITISATION OF INTERNAL CLIENTS

Previous sections described how AUSPEX calculates *compromise* and *exposure* indicators associated with internal clients of the organisation. In this section, we discuss how these indicators are combined to produce the final ranking that is presented to the security analyst.

In this section, we will examine the results of AUSPEX applied to the network traffic generated within a large organisation with 6,432 internal clients. As a first step, AUSPEX calculates the compromise indicators on all internal clients by analysing the network security logs. The output is a set of three compromise APT indicators related to maintaining access, lateral movement, and data exfiltration, denoted by c_1^h , c_2^h and c_3^h. Each index is characterised by different values and ranges, hence the evaluation of an overall *compromise score* C_h for each internal client h requires a normalisation step. For this, AUSPEX uses the *two-sided Quartile Weighted Median* (QWM) metric (Duffield & Lo Presti, 2009). Figure 4 shows an example of how the QWM normalises the distribution of the compromise indicators. The left chart shows a boxplot for each of the compromise indicators. This representation highlights the differences in scale and distribution of the three data sets. The chart on the right shows the same three indicators after normalisation through QWM, and demonstrates how scales and distributions are now comparable. The overall compromise score C_h for each host is calculated as the weighted sum of the three normalised compromise indicators.

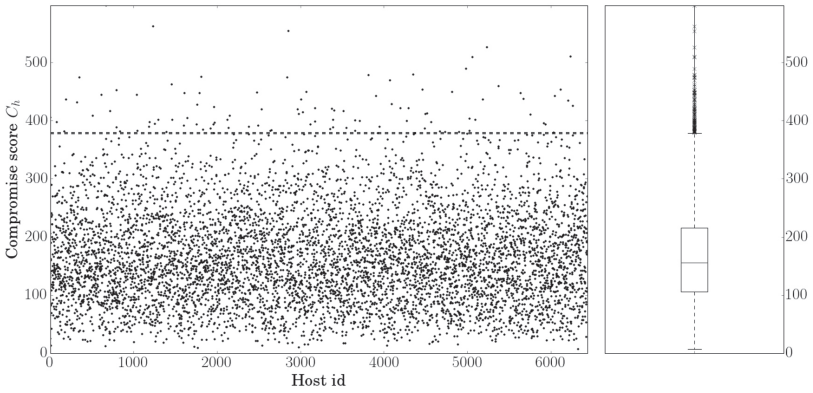
FIGURE 4: EFFECTS OF THE NORMALISATION OF THE COMPROMISE INDICATORS THROUGH THE QWM METRIC



The top-K clients are dynamically determined by applying the *boxplot rule* (Soong, 2004) to identify the hosts characterised by outlier compromise scores. If the compromise score values distribution does not have statistical outliers, then AUSPEX considers the clients whose score

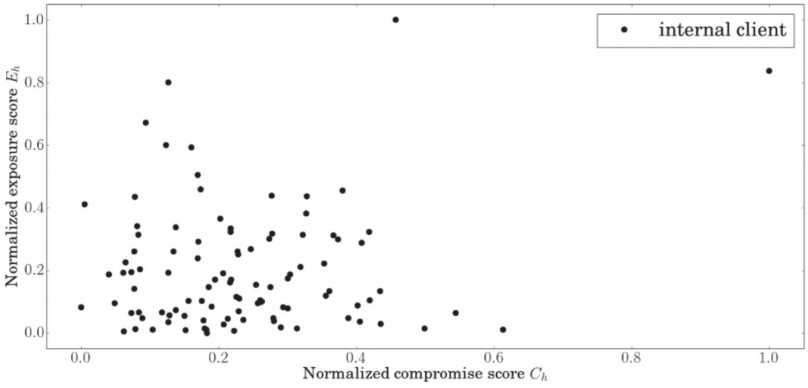
is higher than the 95-th percentile. An example of the choice of the top-K likely compromised internal clients is shown in Figure 5. On the left side, we have a scatterplot where the X-axis represents a *client id* (out of 6,432 clients in the organisation), and the Y-axis is the compromise score C_h. On the right side, we have a boxplot representation of the same distribution. The dashed horizontal line represents the threshold of the boxplot rule (Soong, 2004). In the considered example, only K=102 internal clients (2% of the original 6,432) present a compromise score higher than the threshold and are selected for further analysis.

FIGURE 5: CHOICE OF THE TOP-K LIKELY COMPROMISED INTERNAL CLIENTS



The top-K clients are then prioritised through the four *exposure indicators* e_h^i described in section 5. For each client h in the top-K list, AUSPEX calculates an overall exposure score E_h by normalising the exposure indicators through QWM, as for the C_h indicators. Figure 6 is a representation of the top-K internal clients where the Y-axis represents the *compromise score* C_h, and the X-axis the *exposure score* E_h.

FIGURE 6: NORMALISED COMPROMISED AND EXPOSURE SCORES OF THE TOP-K INTERNAL CLIENTS



To support the security analyst in prioritising analysis of likely compromised internal clients, AUSPEX produces two visual outputs: stacked histograms, and internal communications graphs.

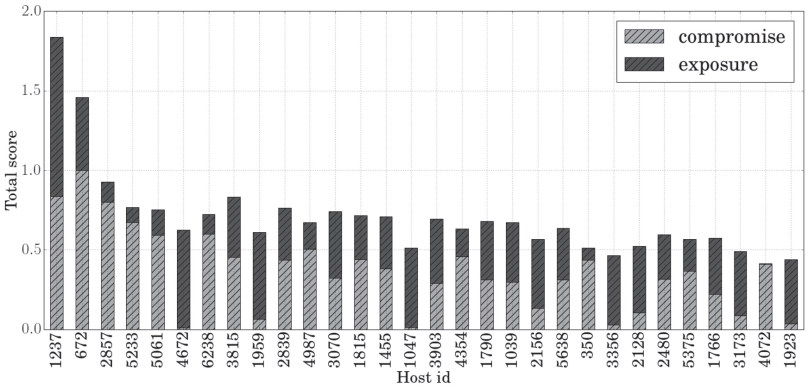
Internal hosts are ranked by considering, for example, the plot of Figure 6 and by computing the Euclidean distance from the origin that is,

$$d_h = \sqrt{E_h^2 + C_h^2}.$$

Results are sorted in decreasing order in Figure 7, where the X-axis shows the client IDs and the Y-axis denotes the sum of the compromise and exposure scores. For the sake of clarity of representation, in Figure 7 we report the internal clients having the thirty highest values of d_h.

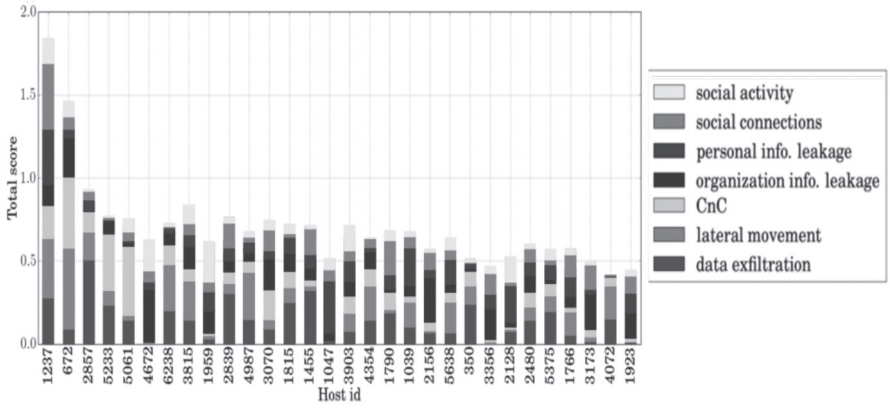
To aid the security analysts, AUSPEX also produces a view including the details of the compromise and exposure indicators. An example is given in Figure 8, where the X-axis shows the client IDs (that are the same of Figure 7) and the Y-axis represents the contribution of each indicator.

FIGURE 7: STACK HISTOGRAM REPRESENTING COMPROMISE AND EXPOSURE SCORES FOR THE TOP-30 INTERNAL CLIENTS



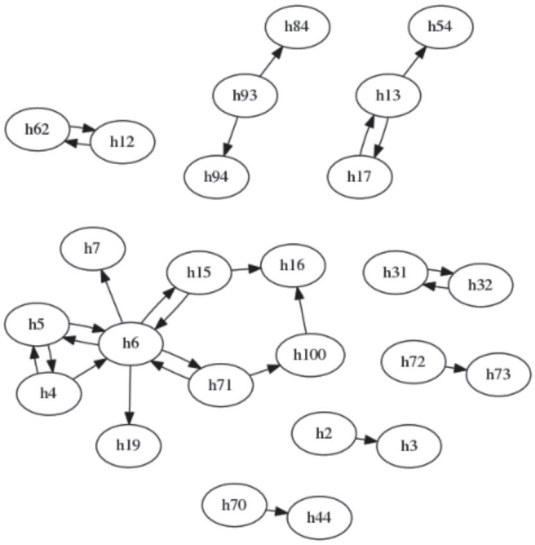
Through AUSPEX it is also possible to show the interactions between the top-K likely compromised clients that may reveal two types of information that is often related to an APT presence: the specific group of hosts that have been hit, and the lateral movements and timing of movements between hosts. These data are fundamental for forensics analyses and incident management processes.

FIGURE 8: STACK HISTOGRAM REPRESENTING THE SCORE DETAILS FOR THE TOP-30 INTERNAL CLIENTS



For these reasons, AUSPEX logs internal communications within the organisation to produce a graph where each node is a top-K likely compromised client, and an edge represents any type of communication between the nodes (e.g., email, chat, or other media used by the organisation). Figure 9 is an example of such a graph, where the node numbers the client order in the ranking considered in Figure 7.

FIGURE 9: EXAMPLE OF GRAPH OF INTERNAL SOCIAL COMMUNICATIONS RELATED TO THE TOP-K LIKELY COMPROMISED CLIENTS



7. RELATED WORK

To the best of our knowledge, AUSPEX is the first framework that ranks the most likely APT compromised hosts of an organisation by combining big data analytics and security intelligence on internal and external information. Big data analytics has already been applied to heterogeneous data to identify security violations (Chari, Habeck, Molloy, Park, & Teiken, 2013), but without a specific focus on APT detection. Other papers (Brewer, 2014; Jeun, Lee, & Won, 2012; Virvilis & Gritzalis, 2013) have analysed the main phases of popular APTs such as Stuxnet, Duqu, and Flame, and limit their proposals to security best practices that could be adopted to prevent them. They do not propose any novel solution or architecture for APT detection.

Other works focus on formalising the APT detection problem and defining possible detection rules, but the chosen approaches, implementation, and testing are left to future work. In this class, we can include several academic papers describing a 7-phase APT detection model (Bhatt, Toshiro Yano, & Gustavsson, 2014; Hutchins, Cloppert, & Amin, 2011); proposing an attack pyramid aiming to capture attacker movements through physical, network and application domains (Giura & Wang, 2012); or suggesting the main building blocks for an APT detection architecture (De Vries, Hoogstraaten, van den Berg, & Daskapan, 2012).

In (Friedberg, Skopik, Settanni, & Fiedler, 2015), the authors propose an anomaly detection system for identifying APTs from security logs. However, their approach requires a huge amount of data to be collected from each host, which is often impractical in large organisations. The analysis and interpretation of its output is also quite difficult and cumbersome because only generic anomalies are identified. Our focus on network logs and open source data makes it more practical, and our output is easier to interpret thanks to the use of several compromise and exposure indicators targeted on specific activities, information and clients.

Other interesting works deal with botnet detection (Bailey, Cooke, Jahanian, Xu, & Karir, 2009). A botnet is a huge set of compromised hosts that are controlled by one or more CnC servers. Several approaches have been proposed for detecting zombies and CnC servers (Gu, Perdisci, Zhang, & Lee, 2008; Gu, Porras, Yegneswaran, Fong, & Lee, 2007), but there are crucial differences that prevent the adoption of botnet detection methods in the APT domain. First, the scale of the problem is completely different, since APTs are human-driven attacks directed at specific organisations and target hosts. Hence, botnet approaches that detect similar behaviours in big groups of hosts (e.g., through clustering of traffic features) are ineffective against APTs that compromise only a few internal hosts. Infection strategies are also different. APTs often use spear phishing and zero-day exploits, while botnets tend to replicate aggressively and automatically (Bailey, Cooke, Jahanian, Xu, & Karir, 2009). AUSPEX is specifically tailored to the APT domain, and takes into account the limitations and challenges that are peculiar to the ranking of internal hosts that perform suspicious network activities.

8. CONCLUSIONS

We design and evaluate a novel framework that is tailored to support security analysts in detecting APTs, which represent the most critical menace to private and public organisations. They are human-driven attacks sustained by complex strategies that combine multidisciplinary skills. Hence, defence approaches based only on automatic methods or on limited information derived by internal sensors do not work because they are affected by too many false positive or negative alarms, respectively. The proposed framework uses multi-factor approaches where big data analytics methods are applied to internal and external information to support (not to replace) human specialists, so that those specialists can focus their security and intelligence analyses on the subset of hosts that are most likely to have been compromised. The proposed approach represents a step forward with respect to the state of the art and paves the way to novel methods for early detection and mitigation of APTs.

REFERENCES

Bailey, M., Cooke, E., Jahanian, F., Xu, Y., & Karir, M. (2009). A survey of botnet and botnet detection. *Conference For Homeland Security, CATCH'09, Cybersecurity Applications \& Technology* (pp. 268-273). IEEE.

Bhatt, P., Toshiro Yano, E., & Gustavsson, P. M. (2014). Towards a Framework to Detect Multi-stage Advanced Persistent Threats Attacks. *IEEE International Symposium on Service Oriented System Engineering (SOSE)* (pp. 390-395). IEEE.

Bilge, L., Balzarotti, D., Robertson, W., Kirda, E., & Kruegel, C. (2012). Disclosure: detecting botnet command and control servers through large-scale netflow analysis. *Proceedings of the 28th ACM Annual Computer Security Applications Conference* (pp. 129-138). ACM.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Brewer, R. (2014). Advanced persistent threats: minimising the damage. *Network Security*, (pp. 5-9).

Brockwell, P. J., & Davis, R. A. (2013). *Time series: theory and methods*. Springer Science & Business Media.

Canali, C., Casolari, S., & Lancellotti, R. (2010). A quantitative methodology to identify relevant users in social networks. *IEEE International Workshop on Business Applications of Social Network Analysis (BASNA)*, (pp. 1-8).

Casolari, S., Tosi, S., & Lo Presti, F. (2012). An adaptive model for online detection of relevant state changes in Internet-based systems. *Performance Evaluation*, (pp. 206-226).

Chari, S., Habeck, T., Molloy, I., Park, Y., & Teiken, W. (2013). A bigData platform for analytics on access control policies and logs. *Proceedings of the 18th ACM symposium on Access control models and technologies (SACMAT '13)*.

Data Breaches. (2016, January). Retrieved from World most popular data breaches, Information is beautiful: http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks.

De Vries, J., Hoogstraaten, H., van den Berg, J., & Daskapan, S. (2012). Systems for Detecting Advanced Persistent Threats: A Development Roadmap Using Intelligent Data Analysis. *IEEE International Conference on Cyber Security (CyberSecurity)*, (pp. 54-61).

Denning, D. E. (1987). An intrusion-detection model. *Software Engineering, IEEE Transactions on*, (pp. 222-232).

Duffield, N. G., & Lo Presti, F. (2009). Multicast inference of packet delay variance at interior network links. *IEEE Computer and Communications Societies*., (pp. 280-285).

Friedberg, I., Skopik, F., Settanni, G., & Fiedler, R. (2015). Combating advanced persistent threats: from network event correlation to incident detection. *Computers & Security*, (pp. 35-57).

Giura, P., & Wang, W. (2012). A context-based detection framework for advanced persistent threats. *IEEE International Conference on Cyber Security (CyberSecurity)*, (pp. 69-74).

Gu, G., Perdisci, R., Zhang, J., & Lee, W. (2008). BotMiner: Clustering Analysis of Network Traffic for Protocol-and Structure-Independent Botnet Detection. *USENIX Security Symposium*, (pp. 139-154).

Gu, G., Porras, P. A., Yegneswaran, V., Fong, M. W., & Lee, W. (2007). BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation. *Usenix Security* (pp. 1-16). Usenix.

Hutchins, E. M., Cloppert, M. J., & Amin, R. M. (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Proceedings of the 6th International Ionference on i-Warfare and Security*.

Irani, D., Webb, S., Pu, C., & Li, K. (2011). Modeling unintended personal-information leakage from multiple online social networks. *IEEE Internet Computing*, (pp. 13-19).

Jeun, I., Lee, Y., & Won, D. (2012). A practical study on advanced persistent threats. *Computer Applications for Security, Control and System Engineering*, (pp. 144-152).

Lam, I.-F., Chen, K.-T., & Chen, L.-J. (2008). Involuntary information leakage in social network services. In *Advances in Information and Computer Security* (pp. 167-183). Springer.

Lindamood, J., Heatherly, R., Kantarcioglu, M., & Thuraisingham, B. (2009). Inferring private information using social network data. *Proceedings of the 18th AMC international conference on World wide web* (pp. 1145-1146). ACM.

Molok, N. N., Chang, S., & Ahmad, A. (2010). Information leakage through online social networking: Opening the doorway for advanced persistent threats. *School of Computer and Information Science, Edith Cowan University, Perth, Western Australia*.

Montangero, M., & Furini, M. (2015). TRank: Ranking Twitter users according to specific topics. *IEEE Consumer Communications and Networking Conference (CCNC)*, (pp. 767-772).

Montgomery, D. C. (1991). *Introduction to statistical quality control*. Wiley New York.

Ostrom, L. T., & Wilhelmsen, C. A. (2012). *Risk assessment: tools, techniques, and their applications*. John Wiley & Sons.

Paxson, V. (1999). Bro: a system for detecting network intruders in real-tim. *Elsevier Computer Networks*, (pp. 2435-2463).

Pierazzi, F., Casolari, S., Colajanni, M., & Marchetti, M. (2016). Exploratory security analytics for anomaly detection. *Computers & Security*, (pp. 28-49).

Romero, D. M., Galuba, W., Asur, S., & Huberman, B. A. (2011). Influence and passivity in social media. *Springer, Machine learning and knowledge discovery in databases*, (pp. 18-33).

Sabahi, F., & Movaghar, A. (2008). Intrusion detection: A survey. *IEEE Systems and Networks Communications (ICSNC)*, (pp. 23-26).

Schiavoni, S., Maggi, F., Cavallaro, L., & Zanero, S. (2014). Phoenix: DGA-based botnet tracking and intelligence. *Detection of Intrusions and Malware, and Vulnerability Assessment (DIVMA)* (pp. 192-211). Springer.

Soong, T. T. (2004). *Fundamentals of probability and statistics for engineers*. John Wiley & Sons.

Virvilis, N., & Gritzalis, D. (2013). The big four-what we did wrong in advanced persistent threat detection? *IEEE International Conference on Availability, Reliability and Security (ARES)*, (pp. 248-254).

Virvilis, N., Serrano, O., & Vanautgaerden, B. (2014). Changing the game: The art of deceiving sophisticated attackers. *IEEE International Conference On Cyber Conflict (CyCon)* (pp. 87-97). IEEE.

Zhou, C. V., Leckie, C., & Karunasekera, S. (2010). A survey of coordinated attacks and collaborative intrusion detection. *Computers & Security*, (pp. 124-140).

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Anonymity Networks and Access to Information During Conflicts: Towards a Distributed Network Organisation

Paolo Palmieri
Department of Computing and Informatics
Bournemouth University
Poole, United Kingdom
ppalmieri@bournemouth.ac.uk

Abstract: Access to information is crucial during conflicts and other critical events such as population uprisings. An increasing number of social interactions happen in the cyberspace, while information exchanges at the infrastructural level (monitoring systems, sensor networks, etc.) are now also based on Internet and wireless links rather than ad hoc, isolated wired networks. However, the nature of the Internet allows powerful hostile actors to block, censor, or redirect communication to and from specific Internet services, through a number of available techniques.

Anonymity networks such as Tor provide a way to circumvent traditional strategies for restricting access to online resources, and make communication harder to trace and identify. Tor, in particular, has been successfully used in past crises to evade censorship and Internet blockades (Egypt in 2011, and Iran in 2012). Anonymity networks can provide essential communication tools during conflicts, allowing information exchanges to be concealed from external observers, anonymised, and made resilient to imposed traffic controls and geographical restrictions. However, the design of networks such as Tor makes them vulnerable to large-scale denial of service attacks, as shown by the DDoS targeted at Tor hidden services in March 2015.

In this paper, we analyse the structural weaknesses of Tor with regard to denial of service attacks, and propose a number of modifications to the structure of the Tor network aimed at improving its resilience to a large coordinated offensive run by a hostile actor in a conflict scenario. In particular, we introduce novel mechanisms that allow relay information to be propagated in a distributed and peer-to-peer manner. This eliminates the need for directory services, and

allows the deployment of Tor-like networks in hostile environments, where centralised control is impossible. The proposed improvements concern the network organisation, but preserve the underlying onion routing mechanism that is at the base of Tor's anonymity.

Keywords: *Tor, anonymous networks, peer-to-peer, denial of service, DDoS*

1. INTRODUCTION

The nature of computer network protocols allows, in principle, a fairly straightforward geographical and organisational mapping of senders and receivers. This can be done both for more restricted local or wireless networks, as well as for the whole Internet. On a large scale, it is thus possible for a government or an Internet service provider to localise, filter, and monitor data streams directed to a specific web service or to a specific geographical region. Several governments effectively control, monitor, or censor Internet traffic, either during crises or permanently. Internet protocols have not been designed for privacy and anonymity, and therefore Internet users can also be easily traced and identified.

This reality has prompted researchers to develop privacy enhancing technologies and anonymity networks, which allow communication to be concealed from external observers, anonymised, and made resilient to control and restrictions. Tor (the 'onion router') is arguably the most successful and widespread anonymity network, counting millions of users [10]. The Tor network is independently developed, and runs on a number of volunteer-operated servers. However, development of the Tor software has been funded by a number of governmental organisations, including US Department of State, DARPA, and the Naval Research Laboratory, as well as the Federal Foreign Office of Germany.[1] This reflects the interest governments around the world have in anonymity networks, which are often seen as both a useful tool and a potential threat [2]. These conflicting sentiments are well exemplified by the discovery in 2007 by security researcher Dan Egerstad that a number of embassies around the world used Tor for delivering private messages, in order not to rely on the hosting country network infrastructure, while the same governments restricted use of Tor by their own citizens [10].

Access to information is critical during conflicts and crises, when controls and restrictions on the flow of information over the Internet are more likely to be imposed [1]. In particular, the ability to use and deploy anonymity networks can be crucial for enabling communication, especially in hostile settings.

A. Onion routing

The Tor anonymity network is built on the concept of *onion routing*. Onion routing was first proposed in 1997 [8], but found widespread use only when implemented in the Tor software in 2002 [5]. The main aim of an onion routing network is to protect users from surveillance and traffic analysis. The identity, location, and network activity of the user are protected by concealing from external observers both the content and routing information of the communication. An onion routing network is therefore a general purpose infrastructure allowing

1 The full list of funders of the Tor project is available on the project's web page: https://www.torproject.org/about/sponsors.html.en [January 4, 2016]

private communications over a public network. It provides anonymous connections that are strongly resistant to both eavesdropping and traffic analysis. This is achieved by relaying the user's traffic, including information about its destination, through virtual *circuits* composed of three successive *relays*. In particular, each relay in the circuit only learns the preceding and following step in the path from the user to the destination: because of this, the user remains anonymous to all relays except the first as well as the destination, while the destination of the communication remains secret to all relays except the last. Messages are also repeatedly encrypted in a layered manner, in order to protect the actual content of the communication: a public key encryption scheme is used to encrypt the communication multiple times, using in inverse order the public keys of all the relays in the circuit. Traffic going back from the destination to the user is similarly encrypted, and routed back from the last relay to the first one.

B. Onion routing during crises

Social media have become an increasingly important mean of communication during crises. In recent years, social media were used extensively in a number of crises, conflicts and uprisings, including the 'April 6 Youth Movement' in Egypt in 2008, the post-election crisis in Iran in 2009, the student protests in Austria in 2009, and the uprisings in Tunisia and Egypt in 2011 and subsequent years part of the larger Arab Spring phenomenon [11]. In all these crises, internet censorship was deployed to prevent access to social media, and anonymity networks played a role in re-enabling access to censored resources, influencing how people and other actors organised online, and ultimately behaved on the streets.

In particular, Tor was used extensively, showing both its strengths and weaknesses. The Tor network, and consequently the onion routing mechanism, proved to be an effective way of circumventing restrictions and internet blockades, while protecting the identity of its users and the secrecy of the communication. However, the open nature of relay servers, which are publicly advertised, make them vulnerable to targeted attacks that can only be partially mitigated by using *bridges* (servers allowing access to the Tor network when direct communication with a relay is impossible). The Tor infrastructure is also limited by the relatively small number of active relays and its semi-centralised structure, which promotes running relays as dedicated servers as opposed to a more distributed, peer-to-peer network organisation [9]. This makes it impossible to use onion routing over local (wireless) networks, which can be potentially deployed on the spot during a crisis using low-cost, low power devices.

Recent events demonstrated that a different implementation of onion routing based on a decentralised network structure might be more suited for crises and conflict areas, where Tor-like networks need to be deployed in hostile environments, and where centralised control is impossible.

C. Onion routing in Wireless Sensor Networks

Another field of application for onion routing is Wireless Sensor Networks (WSN). A WSN is composed of a number of sensors, communicating with each other through a wireless channel, and deployed in an environment where they observe physical phenomena [6]. They are being used in a number of military application scenarios, for purposes including monitoring

and intelligence. WSN are therefore often deployed in hostile or difficult settings, such as battlefields or conflict areas, and are therefore required to be highly fault tolerant, scalable and decentralised. Because of this, WSN are increasingly designed around a distributed network structure and peer-to-peer primitives, to enhance scalability and resilience [7].

Particularly in hostile settings, the security of the communication within the WSN is of paramount importance. Base stations, which are special network nodes that collect data gathered by the other sensors nodes in the WSN, are a central point of failure. It is therefore crucial to make them hard to distinguish from regular nodes. This can be achieved by hiding information on their location and identity (known as context information) within the network [3]. Context information can be protected by employing anonymous routing, and encrypting the communication. In particular, onion routing can be used in a WSN to prevent adversaries from learning the network topology using traffic analysis, and therefore preserve context privacy [4]. However, this requires protocol and mechanisms allowing the deployment of onion routing over the decentralised, peer-to-peer network structures at the base of current WSN.

D. Outline of the paper

The paper is organised as follows. In Section 2.A, we present the challenges of implementing a full onion routing mechanism in distributed networks, and in particular the fact that nodes in a network have limited visibility of the network itself. In Section 2.B, we introduce a novel data structure, called *visibility filter* that enables the sharing of information regarding node visibility across the network in a secure and distributed manner. We present the strategy used to distribute the filters in Section 2.C. Based on the visibility filter structure, we propose an onion routing circuit selection mechanism in Section 2.D. Finally, in Section 2.E we analyse the security of the proposed construction, and in Section 2.F we discuss the communication overhead of the scheme.

2. ONION ROUTING OVER DISTRIBUTED NETWORKS

In the Tor network, clients learn about the currently available relays by downloading the list of running relays from *directory authorities*. Directory authorities are a small subset of relays that collect and distribute routing information in the network [5]. This information is used by the clients when building a circuit, in order to decide which relays to select. Directory information, however, also allows any party (including an attacker) to learn the complete list of relays. The Tor network is based on distributed trust: as the network is open (in the sense that anybody with a sufficiently fast Internet connection can run a relay), it should be hard for a single person or organisation to control large parts of the network. For this reason, directory authorities are selected among long-running, established network nodes. Similarly, the first hop in any circuit built over the Tor network is restricted to being selected among the list of *entry guards* [5]. Both flags (directory authority and entry guard) can be earned by relays after a certain time of continuous operation, proving their stability. This design serves two main purposes: reducing the risk of end-to-end correlation for any given circuit, that is, the chance that both the first and last hop in a circuit are controlled by an adversary; and raising the start-up cost for the adversary.

Without entry guards, the attacker could introduce relays into the network and immediately start having chances to act as first hop. With entry guards, new adversarial relays need to earn the guard flag before they can act as first hop, and the limited number of selected entry guards can prevent attackers from gaining a guard flag for a significant number of relays [12],[13].

While the directory structure and entry guards help protect the privacy of the users, they also expose the Tor network to (distributed) denial of service (DDoS) attacks. As the list of relays and their role is publicly available, an attacker with sufficient resources can target enough relays to entirely disrupt network operation [15], as shown by the large scale DDoS attack targeted at Tor hidden services in March 2015. Worse still, instead of a blanket denial of service attack, an adversary may decide to selectively target relays that are not compromised (or under their control) in order to redirect users to systems one has access to, thus increasing the probability of compromising anonymity [14]. While detection of denial of service attacks is possible [16], and some basic countermeasures such as using client puzzles to mitigate their effect on the network exist [17], DDoS still pose a major threat to Tor, and any other onion routing network based on a similar semi-centralised structure.

A. Challenges of distributed networks

In this paper, we propose a general mechanism for achieving onion routing over a distributed and decentralised network, including network structures organised according to a peer to peer paradigm. Peer to peer (P2P) networks are in fact large decentralised and distributed systems. The design of distributed networks generally follows three main principles: *decentralisation*, in that peers operate without any central coordination; *fault tolerance*, or being able to accommodate nodes joining, leaving, or failing during network operation without a disruption of service; and *scalability*, the property of functioning efficiently for any given number of nodes. Distributed networks are the most efficient and reliable network organisation where network access is limited or restricted. They can allow local nodes to connect to each other, and can be deployed in hostile environments where centralised control is impossible. They are also inherently better suited to coping with denial of service attacks. For this reason, they can be effectively used during conflicts and crises, whether for dedicated networks such as wireless sensor networks or local wireless communication, or overlay networks providing a secure layer over an insecure or openly adversarial network, including the Internet.

However, onion routing cannot be directly implemented over distributed and peer to peer networks without modification. In fact, onion routing makes two important assumptions about the organisation of the underlying network: first, it is assumed that all relays are able to communicate with each other; and second, the Tor directory structure requires that a list of all relays active across the network can be created and maintained. Neither of those assumptions can be satisfied in a distributed network. In a peer to peer setting, nodes do not generally have a full view of the network; that is, they only know a subset of other nodes in the network, called *neighbours*. In fact, nodes are often unable to connect to most other nodes due, for instance, to NATs, firewalls, or, in the case of wireless networks, signal reach). The reduced network visibility means that no node or set of nodes can create with certainty a list of all nodes participating in the network at any given time. This prevents the creation of a directory

structure, informing nodes of the potential relays with which to create a circuit. However, while in Tor the roles of client and relay are generally distinct, it is possible to assume that all nodes in the distributed networks can act as relay. In this setting, nodes can create circuits where the first relay is one of their neighbours. However, this poses a second problem: how can we select following relays, considering that the client node has no knowledge of the network outside its neighbours? This is especially important, as we want to avoid circuits that are too local – that is, that are entirely comprised of neighbours of the client node – in order to avoid partitioning the network. At the same time, we cannot trust relays to select the following steps in the circuit, as doing that would mean that a compromised first hop would be able to influence the creation of the whole circuit, and include only relays under the control of the attacker, thus breaking the user's security.

In order to address this issue, we introduce a mechanism based on controlled flooding and a Bloom filter based data structure, which allows the distribution of relay information among nodes (see Section 2.B). The Bloom filter structure, which we call *visibility filter*, stores information on the neighbours of each relay (and therefore node) in a privacy-preserving way. This information is spread across the local portion of the network through a depth-limited flooding mechanism, where each node transmits its own filter to the neighbours, and at the same time relays the filters received from the neighbours for a limited number of hops, corresponding to the expected length of the circuit (generally 3). The visibility filter structure addresses both assumptions necessary to achieve onion routing: first, it allows relays to learn which other relays they can build a circuit with, without having to learn the relay identities; and second, it supersedes the directory structure and therefore the necessity to compile a list of all relays. The mechanism is introduced in the following section.

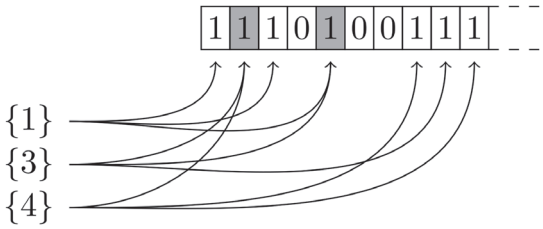
Several attempts have been made to combine mix-net based anonymity networks and P2P. Morphix [22], proposed in 2002, is a peer-to-peer based anonymity network that provides a collusion detection mechanism against an adversary running multiple nodes. However, since the mechanism is based solely on a node's local knowledge, the collusion detection mechanism can be broken, rendering the network insecure [23]. ShadowWalker, proposed in 2009, is a P2P anonymous communication system that aims at being low-latency in nature [24]. However, ShadowWalker circuits are constructed following a traditional random walk strategy that does not address the problem of making sure that not all nodes are in close proximity to the originating peer. Other anonymous routing protocols have been designed for specific P2P network topologies. Salsa [21], NISAN [25] and Torks [26] are all based on the common Distributed Hash Table (DHT) topology. However, the lack of anonymity in their lookup mechanisms may enable an adversary to infer the path structure [27]. In this paper, we propose a solution for implementing onion routing on a peer to peer network, independently of the network topology and structure. Our strategy doesn't require risky network lookups, and ensures that the circuits are not local to the originating node.

B. Bloom filters and visibility filters

A Bloom filter (BF) is a space-efficient data structure representing a set [18]. A BF generated for a set allows the determination, without knowledge of the set itself, of whether an element is in

the set or not, with a probability of false positives p. A Bloom filter can be represented as a binary string of length n, initially all set to 0, and a set of hash functions whose outputs are uniformly chosen in $\{1, \ldots, n\}$. During the creation of the filter, all the elements in the originating set are given as input to each hash function recursively, and bits in the filter corresponding to the output of each hash are set to 1; that is, if for instance one of the hash functions returns the value 5 for an element of the set, the 5th bit of the Bloom filter string is set to 1. A Bloom filter can be queried in the same way; we determine whether an element is part of the originating set by passing its value to the hash functions and reading the bits corresponding to their outputs. If one or more bits have value 0, the element is not part of the set. If instead all bits have value 1, the element is part of the set, minus a false positive probability p. A false positive happens when all the values have been set to 1 during the filter creation by other elements (an event called collision), and not by the element of the query.

FIGURE 1: BLOOM FILTER CONSTRUCTION. IN THE PICTURE, THREE ELEMENTS {1,3,4} ARE ENCODED IN THE FILTER. THE FILTER HAS LENGTH n=10, AND THREE HASH FUNCTIONS ARE USED



We can construct a Bloom filter for each node, containing information on the neighbours of the node itself. The filter can be then used by a third node (that is, a node that is not a neighbour of the first node) to determine whether there is a 'route' between the two nodes: in other words, whether there is a node that is a neighbour of both nodes, allowing them to communicate. This filter, which we call *visibility filter*, can be used when building an onion routing circuit in order to verify that all hops in the circuits are able to communicate with the previous and next hop.

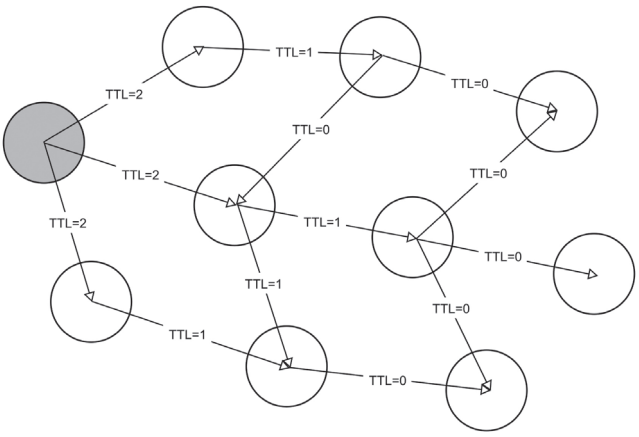
In practice, we assume that each node possesses an identifier unique over the network. The filter is then built over the set of all identifiers corresponding to the neighbours of the node for which the filter is being built. A node can verify the existence of a common neighbour with a node for which it has the visibility filter following a two-steps procedure. First, the node performs a XOR operation over the received filter and its own. If the resulting intersection filter has less than n bits of value 1, then no common neighbour is possible for the properties of Bloom filters. If the number of bits with value 1 is greater or equal to n, then the node proceeds to test each of its neighbours against the other node's filter. Neighbours satisfying the filter are common neighbours, minus a false positive probability p.

C. Distribution of visibility filters

In order to maintain their effectiveness in a constantly changing network, visibility filters need to be recreated and distributed at regular time intervals. We propose to distribute filters across the network using a controlled network flooding mechanism. Following this strategy, each node transmits its own visibility filter to all its neighbours. The identifier of the node is appended to the filter to allow identification of the filter origin (to guarantee integrity and prevent filter forging, the filter can be also signed using the node key pair), as well as a counter flag, called TTL (Time To Live) with value equal to the number of relays in an onion circuit (which defaults to 3). The neighbours will decrease the TTL flag by one, and further forward it to their own neighbours. This process is repeated by all nodes receiving the filter until the TTL reaches 0 (see Figure 2).

While this process involves a communication overhead on the network for filter distribution, it is important to notice that each node benefits from learning as many filters (and therefore potential relays) as possible, in order to enhance the variety of its relay list and therefore mitigate risks posed by adversarial relays. At the same time, limiting the distribution of the filters to the nodes within reach for circuit building reduces the network overhead. We discuss the network overhead in detail in Section 2.F.

FIGURE 2: A SCHEMATIC OF THE LIMITED FLOODING MECHANISM USED FOR DISTRIBUTING THE VISIBILITY FILTER. THE TTL VALUE LIMITS THE DISTRIBUTION OF THE FILTER ONLY TO THOSE NODES THAT CAN BE INCLUDED IN A CIRCUIT



D. Onion circuit creation

Once a node obtains a sufficient number of filters, it can start creating onion circuits for communication. A circuit is created using a trial and error strategy. The ‘client’ is the node (either a user’s device or an autonomous system such as a sensor) which wants to communicate in a secure and private manner; it is, in fact, a client connecting to the service provided by the onion routing network. Communication originating from the client should reach an intended

destination over an onion circuit, of which the client knows the identifier and visibility filter. We assume that the destination is a node of the network. The circuit is built as follows:

1. The client selects a first potential relay for the circuit among its neighbours.
2. Then, the client selects all filters received from neighbours other than the selected first relay. This ensures that the circuit will not be one artificially suggested by the relay itself through manipulation of the filter distribution.
3. Among the eligible filters, the client selects those that satisfy the identifier of the destination node as potential last relays. This step ensures that the eventual last relay will be able to communicate with the destination.
4. The client selects a random relay among the potential last relays. Then, it calculates the intersection between the filters of the potential first and last relay (using a XOR operation): if the number of resulting bits with value 1 is greater than n, a common neighbour exists (minus false probability p). Otherwise, the client selects a different relay and repeats the last step of the process. In practice, this step ensures that the first and last relay will be both able to communicate with at least one common third relay, which will be the middle relay of the circuit.
5. Once two compatible first and last relays are found, the client instructs the first relay to try to build a circuit by sending to it the intersection filter that will be used to identify the common middle relay.
6. The first relay builds the first part of the circuit by connecting to one of its neighbours which satisfies the received intersection filter (the middle relay).
7. Then, the client uses the incomplete circuit to communicate with the middle relay, and instructs it to complete the construction of the circuit by connecting to the last relay.

The circuit creation process enables nodes in a distributed network to reliably build onion circuits without relying on a centralised directory structure. This is made possible by the combined used of visibility filters and a decentralised mechanism for the distribution of relay information.

E. Sybil attacks and security considerations

In general, all distributed and peer to peer networks are vulnerable to Sybil attacks, where the adversary generates malicious nodes in the network in such a way that the target node in most network communication is in some way dependent on them [19]. Sybil attacks are the computer network equivalent of a siege; attack targets are generally surrounded by malicious entities, with the notable difference that there is generally no way to distinguish malicious nodes from honest ones, thus making detection of Sybil attacks more difficult. The general strategy for mitigating the effects of a Sybil attack is to limit the reliance of nodes on their neighbour for communication; if the attacker needs to control or deploy nodes across the whole network, the cost of the attack consequently increases [20].

The circuit creation protocol we propose in this paper achieves reduced locality of the circuit, thanks to the selection of potential relays; neighbours are excluded from acting as second

or third relay, and third relays are selected from those whose information was not originally received from the selected first relay (thus preventing the creation of circuits influenced by a single node). While perfect security against Sybil attacks is generally impossible to achieve, these measures mitigate the impact of such attacks, and therefore increase the privacy and security of the user.

The circuit building mechanism also uses an intelligent selection strategy for relays following the first in order to minimise the impact of a malicious first relay. In fact, at step 2, the client excludes from the set of potential second relays those nodes whose filter was received from the selected first relay. This ensures that a malicious first relay will not be able to influence the selection of the following nodes in the circuit.

F. Performance considerations

In terms of network performance and overhead, the main deviation of the proposed scheme from the classical onion routing implementation is the additional requirement of distributing the visibility filters among nodes. In this Section, we describe why the limited flooding strategy we propose is realistic and introduces only limited overhead.

In general, peer-to-peer networks adopt different strategies for distributing or searching for information among peers. The concept of *flooding* was introduced by the Gnutella 2000 search protocol [28]. In practice, a node looking for a specific resource over the network broadcasts its query to its neighbours. If a neighbour does not have the resource, it forwards the query to its own neighbours. This is repeated until the resource is found, or all the nodes have been contacted. This naïve approach scales very poorly in large networks. For this reason, several alternative approaches have been proposed and implemented in subsequent networks that modify the flooding behaviour [29]. The main issue with network flooding is the high network use it can generate. This, combined with the uncertainty of the timing of queries (especially when user generated) can result in a significant overhead. In the proposed scheme, we address this by adopting a smart flooding strategy that limits the impact of the visibility filter distribution over the network. In particular, we limit the depth of the flooding (that is, the number of times a filter is relayed). This restricts the communication to a small portion of the network, and consequently greatly reduces network use. We can safely do so because we know exactly how far the information should be transmitted: as relays will only communicate with neighbouring relays, the distance coincides with the length of a circuit. At the same time, we can control the timing of the flooding, thus preventing network overload. In fact, we can define regular interval at which the nodes in the network should transmit their filters, and can decide to limit this further by only transmitting filters if there is a significant change in the neighbour's set. New nodes entering the network can request a cached set of filters from their neighbouring nodes, which can easily keep them until a new transmission due to the very limited space requirements of Bloom filters.

We can estimate the network use of the limited flooding strategy. If we assume that each node has 100 neighbours, the size of the filer will be 839 bytes (for a false positive probability of less than one in a million, or e^{-14}). If the percentage of common neighbours between two

nodes is 20%, the overall network use for a single node participating in the distribution of the filter can be estimated to approximately 83KB. Please note, this is the size of the information being transferred (the payload): the actual network use depends on implementation details (and in particular, on the chosen communication protocol and its parameters). This overhead is perfectly compatible even with networks with limited bandwidth, such as WSNs.

3. CONCLUSIONS

In this paper, we discussed how a distributed network organisation could overcome many of the limitations and security challenges posed to networks deployed in difficult or hostile settings, such as during crises and conflicts. We analysed how onion routing can be used to secure communication under those circumstances. We also presented the limitations that the original design of onion routing and its most common implementation, Tor, have when used on distributed networks. In order to address these shortcomings, we proposed a number of modifications to the onion routing mechanism aimed at improving its resilience to a large coordinated offensive run by a hostile actor in a conflict scenario, such as a denial of service attack. In particular, we introduce novel mechanisms that allow relay information to be propagated across the network without requiring a directory structure, and we address the issue of the limited visibility among nodes of distributed networks. These results open the way to the deployment of onion-routing-enabled networks in hostile environments, where centralised control is impossible.

A natural direction for future research would be an implementation of the proposed scheme, to provide experimental results on the overhead incurred by the distributed network for the exchange of the relay visibility information.

ACKNOWLEDGMENTS

The author would like to thank Johan Pouwelse for interesting discussions on the topic of the paper, as well as for previous collaborations on peer-to-peer networks.

REFERENCES

[1] Mina Rady, Nazli Choucri. ‘Anonymity Networks: New Platforms for Conflict and Contention’. Available: http://web.mit.edu/minarady/www/papers/Tor Conflict -Mina Rady-v3.0 single .pdf (January 4, 2016).

[2] Emin Çalışkan, Tomáš Minárik, Anna-Maria Osula. ‘Technical and Legal Overview of the Tor Anonymity Network’. NATO Cooperative Cyber Defence Centre of Excellence. Available: https://ccdcoe.org/sites/default/files/multimedia/pdf/TOR_Anonymity_Network.pdf (January 4, 2016).

[3] Li, N., Zhang, N., Das, S.K., Thuraisingham, B.M. ‘Privacy preservation in wireless sensor networks: A state-of-the-art survey’. Ad Hoc Networks, vol. 7(8), pp.1501-1514, 2009.

[4] Paolo Palmieri. ‘Preserving Context Privacy in Distributed Hash Table Wireless Sensor Networks.’ In Information and Communications Security - 17th International Conference (ICICS 2015), Revised Selected Papers. Lecture Notes in Computer Science, Springer, vol. 9543, pp.436-444, 2016.

[5] Roger Dingledine, Nick Mathewson, Paul Syverson. ‘Tor: The Second-Generation Onion Router’. In Proceedings of the 13th Usenix Security Symposium (Usenix 2004), pp.303-320, 2004.

[6] Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E. 'Wireless sensor networks: a survey'. Computer Networks, vol. 38(4), pp.393-422, 2002.
[7] McGoldrick, C., Clear, M., Carbajo, R.S., Fritsche, K., Huggard, M.. 'Tiny Torrents: Integrating peer-to-peer and wireless sensor networks'. In the Proceedings of the Sixth International Conference on Wireless On-Demand Network Systems and Services (WONS'09), IEEE Press, pp.109-116, 2009.
[8] Syverson, P.F., Goldschlag, D.M., Reed, M.G. 'Anonymous connections and onion routing.' In the Proceedings of the 1997 IEEE Symposium on Security and Privacy, IEEE Computer Society, pp.44-54. 1997.
[9] Paolo Palmieri, Johan Pouwelse. 'Key Management for Onion Routing in a True Peer to Peer Setting.' In Advances in Information and Computer Security - 9th International Workshop on Security (IWSEC 2014), Proceedings. Lecture Notes in Computer Science, vol. 8639, pp.62-71, Springer, 2014.
[10] Damon McCoy, Kevin Bauer, Dirk Grunwald, Tadayoshi Kohno, Douglas Sicker. 'Shining Light in Dark Places: Understanding the Tor Network.' In the Proceedings of the Eighth International Symposium on Privacy Enhancing Technologies (PETS 2008). pp 63-76, 2008.
[11] Stefan Csizmazia. 'The Role of Online Social Networks in Political Uprisings.' Available: http://www.steviec.at/uni/bakk/csizmazia__role_of_osn_in_political_uprisings.pdf (January 4, 2016).
[12] Karsten Loesing. 'Measuring the Tor Network from Public Directory Information.' In the Proceedings of the 2nd Hot Topics in Privacy Enhancing Technologies (HotPets 2009), 2009.
[13] Kevin Bauer, Damon McCoy, Dirk Grunwald, Tadayoshi Kohno, Douglas Sicker. 'Low-Resource Routing Attacks Against Tor.' In the Proceedings of the 2007 ACM workshop on Privacy in electronic society (WPES 2007), ACM, pp 11-20, 2007.
[14] Nikita Borisov, George Danezis, Prateek Mittal, Parisa Tabris. 'Denial of service or denial of security?' In Proceedings of the 14th ACM conference on Computer and communications security (CCS '07), ACM, pp.92-102.
[15] Norman Danner, Sam Defabbia-Kane, Danny Krisanc, Marc Liberatore. 'Effectiveness and detection of denial-of-service attacks in Tor'. ACM Transactions on Information and System Security, vol. 15(3), article 11, 25 pages, November 2012.
[16] Norman Danner, Danny Krisanc, Marc Liberatore. 'Detecting Denial of Service Attacks in Tor'. In the Proceedings of Financial Cryptography and Data Security, 13th International Conference (FC 2009), Lecture Notes in Computer Science, vol. 5628, pp.273-284, Springer, 2009.
[17] Fraser, N.A.; Kelly, D.J.; Raines, R.A.; Baldwin, R.O.; Mullins, B.E., 'Using Client Puzzles to Mitigate Distributed Denial of Service Attacks in the Tor Anonymous Routing Environment'. In the Proceedings of the 2007 IEEE International Conference on Communications (ICC 2007), pp.1197-1202, 2007.
[18] B.H. Bloom. 'Space/time trade-offs in hash coding with allowable errors'. In Communications to the ACM, vol 13(7), pp.422-426, 1970.
[19] John R. Douceur. 'The Sybil Attack'. In the Proceedings of 1st International Workshop on Peer-to-Peer Systems (IPTPS), pp.251-260, 2002.
[20] Rowaihy, H.; Enck, W.; McDaniel, P.; La Porta, T., 'Limiting Sybil Attacks in Structured P2P Networks.' In the Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM 2007). IEEE, pp.2596-2600, 2007.
[21] Arjun Nambiar, Matthew Wright. 'Salsa: a structured approach to large-scale anonymity'. In Proceedings of the 13th ACM conference on Computer and communications security (CCS '06), ACM, 2006.
[22] Marc Rennhard, Bernhard Plattner. 'Introducing MorphMix: peer-to-peer based anonymous Internet use with collusion detection.' In Proceedings of the 2002 ACM workshop on Privacy in the Electronic Society (WPES '02), ACM, 2002.
[23] Parisa Tabris, Nikita Borisov. 'Breaking the Collusion Detection Mechanism of MorphMix'. In Proceedings of the 6th International Workshop on Privacy Enhancing Technologies (PET 2006), Lecture Notes in Computer Science, vol. 4258, pp.368-383, Springer, 2006.
[24] Prateek Mittal, Nikita Borisov. 'ShadowWalker: peer-to-peer anonymous communication using redundant structured topologies'. In Proceedings of the 16th ACM conference on Computer and communications security (CCS '09), ACM, 2009.
[25] Andriy Panchenko, Stefan Richter, Arne Rache. 'NISAN: network information service for anonymisation networks.' In Proceedings of the 16th ACM conference on Computer and communications security (CCS '09), ACM, 2009.
[26] Jon McLachlan, Andrew Tran, Nicholas Hopper, Yongdae Kim. 'Scalable onion routing with torsk', In Proceedings of the 16th ACM conference on Computer and communications security (CCS '09). ACM, 2009.

[27] Qiyan Wang, Prateek Mittal, and Nikita Borisov. ‘In search of an anonymous and secure lookup: attacks on structured peer-to-peer anonymous communication systems’, In Proceedings of the 17th ACM conference on Computer and communications security (CCS ‘10), ACM, 2010.
[28] M. Ripeanu, ‘Peer-to-peer architecture case study: Gnutella network.’ In Proceedings of the First International Conference on Peer-to-Peer Computing, 2001, pp.99–100, 2001.
[29] Niels Zeilemaker, Zekeriya Erkin, Paolo Palmieri, Johan A. Pouwelse. ‘Building a privacy-preserving semantic overlay for Peer-to-Peer networks’. In 2013 IEEE International Workshop on Information Forensics and Security (WIFS 2013), pp.79-84, IEEE, 2013.

2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

We Know Where You Are!

Siddharth Prakash Rao
Department of Computer Science
Aalto University, Finland
siddharth.rao@aalto.fi

Dr Silke Holtmanns
Bell Labs, Nokia
Espoo, Finland
silke.holtmanns@nokia.com

Dr Ian Oliver
Bell Labs, Nokia
Espoo, Finland
ian.oliver@nokia.com

Dr Tuomas Aura
Department of Computer Science
Aalto University, Finland
tuomas.aura@aalto.fi

Abstract: Mobile network technologies require some degree of tracking of user location, specifically user equipment tracking, as part of their fundamental mechanism of working. Without this basic function, features such as hand-over between cells would not work. Since mobile devices are typically associated with a single person, this provides a potential mechanism for user location surveillance. Network operators are bound by strict privacy legislation. However, spying by certain agencies, hackers and even advertisers without the users' or operators' knowledge has become a serious issue. In this article, we introduce and explain all known recent attacks on mobile networks that compromised user privacy. We focus on attacks using the Signalling System 7 (SS7) protocol as the interconnection interface between operators mainly in GSM networks. In addition, we outline a novel evolution of location tracking for LTE networks. One reason these attacks are not widely published or known by the general public is due to the complex and arcane nature of the networks and their protocols. Mobile network interfaces are 'hidden' from users, and therefore the general public's interest in such attacks is much lower compared to other phone vulnerabilities.

The purpose of the paper is to raise awareness about the current location tracking problem in cellular networks, the existing countermeasures and to encourage further research in the area for 5G networks.

Keywords: *location privacy, SS7, mobile networks, interworking, roaming, tracking, diameter*

1. INTRODUCTION

Mobile phones have become a major part of our daily lives and communication. Mobile phone services are based on cellular network technology which requires the operators to keep track of users' movements between cells and networks in order to provide seamless telecommunications services such as network access, messaging, and voice calls direct to the phone. Since personal mobile phone use is nearly ubiquitous, the discourse of location (i.e. Cell ID) information that is collected by the network operators poses a threat to the personal privacy. Disclosure of location information by the operators is strictly controlled by legislation in most countries. However, spying on mobile users by government agencies, hackers, and advertisers without the knowledge of the user or the network operator has become a serious issue.

Cellular location privacy research is related to the disclosure of identifiers in the Radio Access Network (RAN) using so-called IMSI catchers with which attackers can spoof base stations and collect mobile subscriber identifiers over the air. This requires the attacker to set up a base station near the assumed location of the target users. Solutions to this problem are being developed to detect and prevent attacks that use a false base station [3]. It is more practical for the attacker to obtain the location information from the cellular network operators directly.

Signalling System No. 7 (SS7) is a widely used protocol for communication between network elements and between operator networks. It is one of the key protocols to enable roaming and cellular services across operator domains. Although there are newer protocols (specifically, Diameter), SS7 is still widely used between cellular networks, and interoperability will force operators to support it long into the future. SS7 was designed when there were a few state owned operators, which in turn trusted each other. The protocol itself offers little or no protection, nor was it designed to resist attacks using the SS7 signalling networks or SIGTRAN (SS7 over IP).

In 2008 the first location tracking attack was illustrated by Engel in [1]. In 2014 it was proven that an attacker with access to the SS7 network can track users and perform other attacks such as eavesdropping, SMS interception, and fraud [1],[2],[4],[5],[6],[22]. One of those attacks was shown in a live demonstration [7].

In this paper, we discuss the known weaknesses in the mobile communication backend of the networks that may disclose the location of a user. We assume that the attacker has the victim's phone number and SS7 access. We provide message-level details of the attacks that exploit the SS7 protocol in addition to outlining a new Diameter based location tracking attack for LTE networks and discussing the potential countermeasures.

2. BACKGROUND

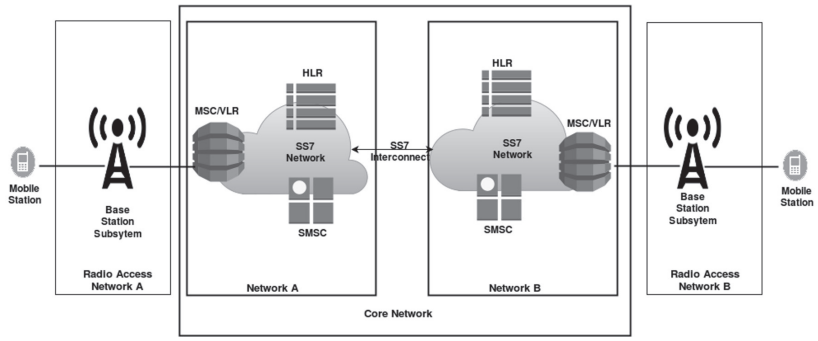
This section gives a brief overview of the SS7 interworking network, the situations where location is intentionally disclosed, and the different levels of accuracy at which the network may reveal the user location.

A. SS7 and interworking network overview

SS7 is a network protocol used worldwide between network elements and between operator networks. It was standardised by the International Telecommunication Union, Telecommunication Standardisation Sector (ITU-T) more than 30 years ago. SS7 specifies the exchange of information over the signalling networks in Public Switched Telephone Networks mainly to enable the establishment of phone calls across networks. The Message Application Protocol (MAP) which is standardised by the 3rd Generation Partnership Project (3GPP) [8] offers a wide set of additional features for enabling mobility, roaming, SMS, and billing.

The Home Location Register (HLR), Mobile Switching Centre (MSC), Visitor Location Register (VLR), and Short Message Service Centre (SMSC) are some of the key components of the core network (shown in Figure 1) used for the attacks discussed here. These elements are identified by their Global Title (GT), which are used as addresses for routing messages through the SS7 network using the SS7 MAP protocol [8] .

FIGURE 1: CORE NETWORK OVERVIEW



HLR is the central database in an operator's home network. It contains the subscription profiles, service data, and current location of the subscribers of that operator. It maintains the mapping of subscribers' International Mobile Subscriber Identity (IMSI) and their phone numbers, or Mobile Station International Subscriber Directory Number (MSISDN). The VLR stores a copy of the data from HLR for mobile subscribers who are currently in its geographic area, for both local and roaming subscribers. The MSC is responsible for routing calls and SMS text messages to and from the mobile phones in the RAN. The SMSC is responsible for storing, forwarding, and delivering SMS messages.

B. Regular and legitimate location disclosure

The radio network to which the user is currently connected knows the precise location of the cell tower and this then gives the approximate location of the user based on proximity measurements and triangulation. In city areas this is up to 200 m around a given cell tower, and in rural areas up to 35 km with normal equipment or up to 77 km with extended cell range equipment. In densely populated areas, more base stations are deployed and each of them covers a small area, which implies more accurate user location.

This information is revealed by some legitimate services as follows:

- Locate-my-phone services. Network operators offer a service for tracking lost phones with the consent of the phone owner. However, these have not gained popularity due to similar functionality provided by GPS.
- Public safety. In case of emergency or when the user has to be tracked down for safety reasons, officials are given access to location information.

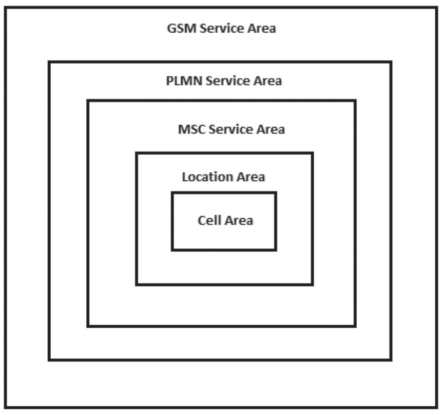
The FCC [9] requires the precise location of the caller for emergency purposes. Access to this level may be protected by local laws and require judicial intervention to obtain. Mobile operators provide the position of the mobile phones as part of the Location Services (LCS). The precise location information is obtained with the Radio Resource LCS Protocol [10] with the help of Gateway Mobile Location Centre (GMLC).

C. Overview of location proximity

To provide cellular services to mobile users, mobile networks have a hierarchical geographic structure [11] (see Figure 2). A cell is the smallest area, and its radius ranges from 100 meters to over 35 kilometres (which is the normal radio coverage of a transmitter). Each cell is identified by the Cell Global Identity (CGI), or Cell ID. When used for positioning, the CGI is typically mapped to the geographic coordinates of the cell tower at its centre.

Several such cells constitute a Location Area (LA). Every time a mobile user moves to a new LA, their location will be updated in the MSC/VLR database. An MSC Service Area comprises several LAs, which are served by the same MSC. HLR stores the information about the MSC that currently serves a particular Mobile Station (MS). Each mobile operator has several MSCs, which together form the PLMN Service area. The overall area with GSM connectivity is called the GSM Service area.

FIGURE 2: GSM GEOGRAPHIC HIERARCHICAL STRUCTURE [11]

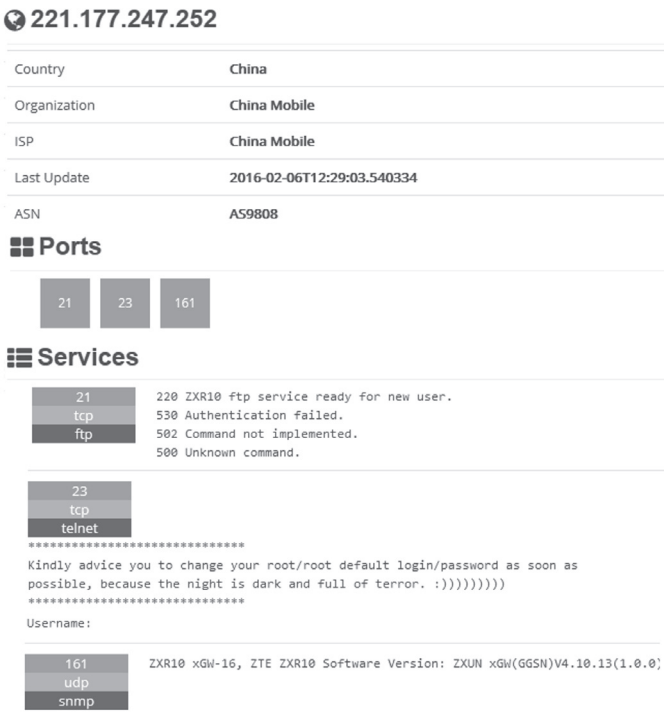


3. LOCATION DISCLOSURE ATTACKS

Attacks from the interconnection network that would reveal the precise location of users were first demonstrated by Engel in 2008 [1] and then, in 2014 [2], more accurately at the Cell ID level. These exploited flaws in the existing specification and implementations. It is possible for an attacker to gain access to the SS7 core network through a compromised edge device or through a badly configured core network node. It might be as easy as searching for a core network node with open ports in an Internet-connected database [12] as shown in Figure 3. Another means of getting into SS7 networks is by gaining connection through an existing provider with insufficient security checks when renting out their SS7 access.

As part of mobility management, network operators have to keep track of the location of the mobile station. This happens even when the mobile is in the idle state, i.e. it is turned on and ready to make or receive messages and calls. The HLR of the mobile phone's home operator needs to know the MSC/VLR via which the mobile can be reached. The MSC/VLR where the mobile is currently roaming needs to be able to page the mobile when a call or message arrives. The mobile phone is identified by its IMSI, which is also used in the mobility management messages between the HLR and MSC/VLR.

FIGURE 3: GGSN VISIBLE ON THE INTERNET (DISCOVERED VIA SHODAN.IO, 11.2.2016)



A. Location disclosure using call setup messages

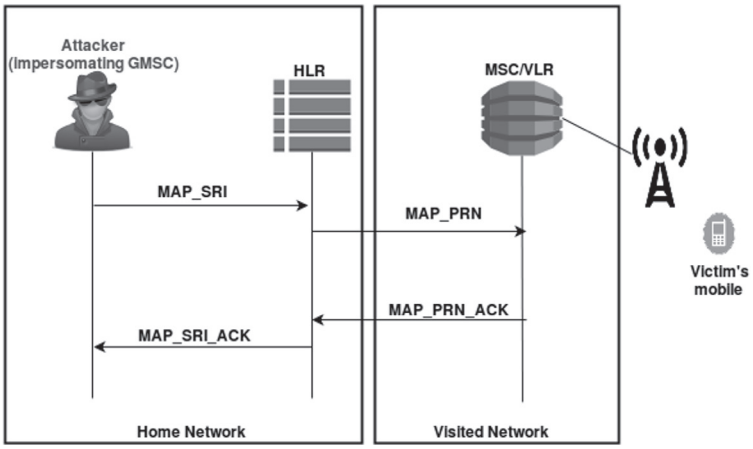
We now describe the normal message flow for a mobile-terminated call when the mobile is roaming in a visited network [13]. The call in this case may originate either from the fixed telephone network or from a mobile subscriber from another operator network.

1. When a call is placed to the mobile user's phone number (MSISDN), the caller's network sends an ISUP IAM (Initial Address Message) message to the mobile user's home network. The message is routed to a Gateway MSC (GMSC) of the mobile's home network based on the MSISDN.
2. The GMSC queries the HLR in the home network for the mobile's current location by sending the MAP Send Routing Information (MAP_SRI) message to the HLR. The HLR keeps track of the mobile's location. We assume that the mobile is roaming in another network.
3. The HLR queries the MSC/VLR in the visited network by sending the MAP Provide Roaming Number message (MAP_PRN).
4. The VLR responds to HLR with the MAP_PRN_ACK message. The response contains the Mobile Station Roaming Number (MSRN), which is a temporary ISDN telephone number assigned by the VLR for the purpose of routing this call.
5. The HLR passes the MSRN back to the GMSC with the MAP Routing Information Acknowledgement MAP RIA message.
6. The GMSC now sends the IAM message to the MSRN to set up the call.
7. This message is routed to the MSC/VLR in the visited network. Since MSC/VLR just assigned the MSRN to the mobile, it knows to which mobile (identified by IMSI and TMSI) the call should be routed. Thus, on arrival of IAM message, it establishes the call connection to the mobile.

Attack using call set up messages. This attack [1] uses the normal message flow of the call set up to learn the approximate location of the victim's phone and therefore of its user. An attacker with SS7 access pretends to be the GMSC and follows the call setup procedure from the point where the GMSC supposedly received the IAM message. The attack message flow, also shown in Figure 4, is as follows:

1. The attacker sends the MAP_SRI message enclosing the victim's MSISDN to the HLR in the victim's home network. In the SS7 network, no authentication is performed. However, the attacker needs to know the Global Title of the HLR to send this message to (but often brute force attacks to operator ranges are performed till an 'HLR is hit').
2. HLR maps the MSISDN to the victim's IMSI and sends MAP_PRN to the VLR at the visited network.
3. The VLR responds with MAP_PRN_ACK, which contains the MSRN (or an error message if the phone is not reachable). The same message contains the IMSI and Global Title of the MSC/VLR that is currently serving the mobile.
4. HLR forwards the information to the attacker, who is impersonating a GMSC, in MAP_SRI_ACK message. This response also contains the Global Title of the VLR/MSC.

FIGURE 4: LOCATION DISCLOSURE USING CALL SETUP MESSAGES



The attacker will not proceed with the call setup to the GMSC. Instead, the attacker has learned the victim's IMSI and the GMSC and Global Title of the VLR of the network where the victim is currently roaming. The latter two disclose the mobile's location with the relatively coarse granularity of the MSC service area.

The MSC service area is typically a region or state of a country. Since the number of mobiles that an MSC can serve is limited, the area served by one MSC is smaller in densely populated areas. The numbering of the MSCs is purely operator specific, but for some operators, the number itself can reveal the geographic area, such as telephone area code of the MSC GT [1]. The GT identifies the country and operator in whose network the mobile is roaming. Another way to discover the location of the GT is to search in a business or residential phone list for phone numbers that have the same prefix as the GT. The addresses listed with these numbers will mostly be in the geographic area of the MSC. The prefix of the MSRN reveals the same kind of information as the GT. The IMSI may be useful for further attacks e.g. fraud, eavesdropping.

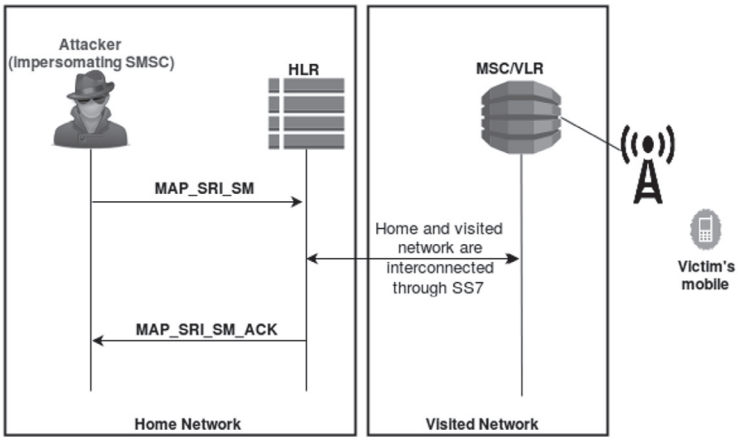
B. Location disclosure using SMS protocol messages

SMS is a service for the transmission of text messages up to 140 bytes. The end-to-end SMS procedure comprises of two parts: in the Mobile Originating Part, the sender submits the SMS to a Short Message Service Centre (SMSC); in the Mobile Terminated Part, the message is delivered from the SMSC to the recipient mobile. The messages are sent over the GSM signalling channels [13].

The message flow of the Mobile Terminated Part is similar to the call setup described above. To deliver the message directly to the destination, the SMSC has to know the IMSI of the recipient mobile and the Global Title of the MSC that is currently serving the recipient. When the SMSC is in the same network as the recipient or a roaming partner, the SMSC obtains the required information from the recipient's HLR:

1. The SMSC sends MAP Send Routing Information for SM MAP_SRI_SM message to the HLR in the recipient's home network.
2. The HLR queries the MSC/VLR in the visited network, where the mobile is currently roaming, with the MAP_PRN.
3. HLR encapsulates the received IMSI and MSC/VLR GT in the MAP Send Routing Information for SM ACK message and sends it back to the SMSC.
4. The SMSC then sends the text message with the IMSI to the recipient MSC/VLR GT, and the MSC/VLR delivers it to the mobile station.

FIGURE 5: LOCATION DISCLOSURE USING SMS PROTOCOL MESSAGES



Attack using SMS protocol messages. Here the attacker impersonates an SMSC and sends signalling messages to learn the IMSI and MSC/VLR GT of the victim [1]. The attack message flow is as follows; it is also shown in Figure 5.

1. Pretending to be an SMSC, the attacker sends the MAP_SRI_SM message to the HLR by enclosing the MSISDN (phone number) of the victim.
2. The HLR thinks that the SMSC needs to send an SMS to the provided MSISDN, and replies with the MAP_SRI_SM_ACK message, which contains the IMSI of the victim along with the GT of the MSC/VLR that is currently serving the victim.

The attacker translates the MSC/VLR GT to the geographic location in the manner described before. However, the success of this attack depends on details of the bilateral SMS roaming arrangement between the network operators.

C. Location disclosure using CAMEL location management function messages

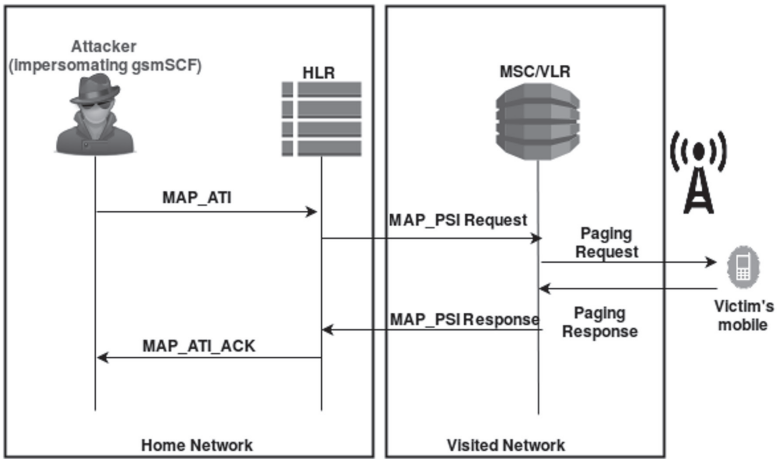
Customised Applications for Mobile Networks Enhanced Logic (CAMEL) [14] is an overlay on MAP. As a part of the network internal location management function, the network providers

can send Any Time Interrogate (ATI) messages to the HLR from CAMEL platforms to obtain the Cell ID of the user. The location information provided in this case is the last known location of the mobile station. The basic message flow of this function [15],[8] is as below:

1. The GSM Service Control Function (gsmSCF) element sends the MAP Any Time Interrogation Request (MAP_ATI) message, which contains the MSISDN, to the HLR of the mobile's home network.
2. The HLR again looks up the mobile's current location in its database based on the MSISDN. It then transmits the MAP Provide Subscriber Information (MAP_PSI) message to the MSC/VLR.
3. The MSC/VLR sends a Paging Request message to the mobile station to look up its current state. The Paging Response message will have the Cell ID of the cellular tower to which the mobile is currently connected.
4. MSC responds to the HLR with the MAP Provide Subscriber Information Response message, which contains the Cell ID and IMSI. If the mobile station responded to the paging, then the age field is set to 0, as the MSC/VLR accurately knows its current location; otherwise the last known Cell ID is sent, with a non-zero age field.
5. The HLR now sends the MAP Anytime Interrogation Response back to the gsmSCF with the subscriber information from the previous step.

Attack using Any Time Interrogation message. Here the attacker impersonates the gsmSCF and sends the MAP_ATI message with the MSISDN of victim to the HLR [B]. The message flow of this attack is shown in Figure 6.

FIGURE 6: ATTACK USING ANY TIME INTERROGATION MESSAGE



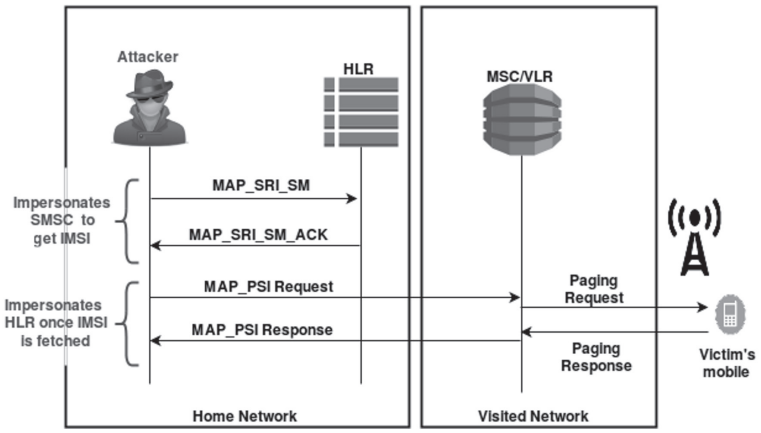
1. The attacker pretends to be a gsmSCF and sends the MAP_ATI request containing the MSISDN of the victim to the HLR.
2. The HLR treats this as a legitimate message from gsmSCF and sends the MAP_PSI request to the MSC/VLR.
3. The MSC/VLR initiates the Paging Request to the victim's phone and receives the IMSI and Cell ID in the Paging Response message.
4. The MSC/VLR then sends the information from the previous step along with the MSC/VLR GT to the HLR in the MAP_PSI response.
5. The HLR forwards this information to the attacker in the MAP_ATI response.

As the result, the attacker now knows the cell of the victim along with the IMSI and the GT of the serving MSC. Here, the attacker would learn victim's location more accurately than in the previous attacks because the Cell ID is retrieved.

Since the MAP Any Time Interrogation message is not an essential function for the network operation and it raises obvious privacy concerns, many network operators filter the message. The filtering can potentially be bypassed by the following hybrid attack.

Hybrid attack using SMS and CAMEL messages. The hybrid attack [6] queries the MSC/VLR directly in order to circumvent the potential MAP_ATI filters. The attacker can send the Provide Subscriber Information request to the MSC/VLR by pretending to be the HLR. For this, the attacker needs to know the victim's IMSI. The attacker needs to discover first the IMSI corresponding to the MSISDN. The message flow for this attack is shown in Figure 7.

FIGURE 7: HYBRID ATTACK USING SMS AND CAMEL MESSAGES



1. The attacker performs the previously described SMS-based attack to learn the IMSI and MSC/VLR GT.
2. The attacker then queries the MSC/VLR with the MAP Provide Subscriber Information MAP_PSI request.
3. As before, after the paging procedure, the MSC/VLR returns the Cell ID to the attacker.

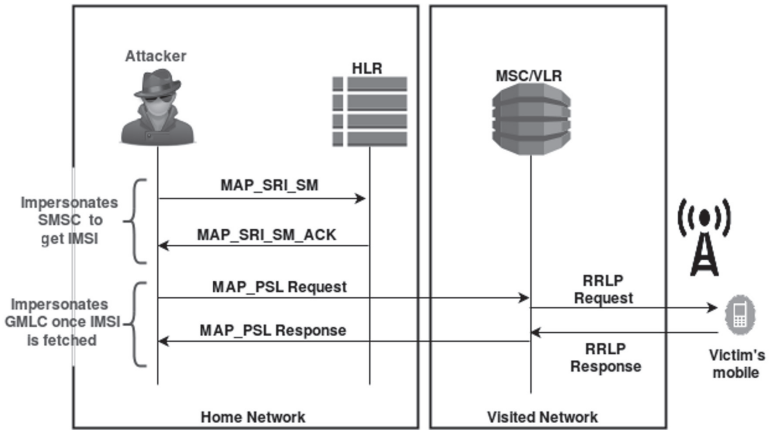
D. Location disclosure emergency location service messages

In situations, governmental bodies have access to location information that is collected from the network [10]. The accurate location is based on triangulation using the angle of signals observed at the cellular towers or the arrival times of the radio signals at the mobile. The triangulation is usually triggered by an emergency call from the mobile station, but it can also be initiated from the network side, such as by law-enforcement officials. The basic message flow of an authorised LCS [16] is as follows:

1. An authorised client can send the MAP Location Service request to a Gateway Mobile Location Centre (GMLC) either at the mobile's home network or at the visited network. The request contains either the MSISDN or the IMSI of the mobile. The authorisation of the client is left to the operator and depends upon local legislation.
2. The GMLC that receives the request sends it to a GMLC in the mobile's home network, which enforces the user's privacy preferences on the request, and then forwards it to the visited network's GMLC. The GMLC may query the HLR at the mobile's home network for the IMSI and routing information with the MAP Send Routing Info for LCS MAP_SRI_LCS.
3. The visited network's GMLC acts as a gateway for external LCS clients to the LCS functions in the radio access network. In 3G networks, the GMLC sends the MAP Provide Subscriber Location MAP_PSL request to the MSC/VLR. This request identifies the mobile station by its IMSI.
4. The MSC/VLR resolves the location request with the help of the radio network and the mobile using one of various positioning methods. It then encapsulates the location report into the MAP Provide Subscriber Location ACK message to the GMLC of the visited network.
5. The GMLC encloses the location information in the MAP Location Service Response and sends it back to SMLC client via the same route as the request was received.

The validation of the request origin and the authorisation are enforced by the GMLCs. Emergency services are allowed to make location requests directly to the GMLC in the visited network. This request is not routed through the GMLC in the home network and can be done without querying the HLR for the IMSI. This allows assistance to users in need also while they are roaming, but brings a way to circumvent control setting in the home GMLC.

FIGURE 8: ATTACK USING LOCATION SERVICE MESSAGES



Attacks using location service (LCS) messages: In this attack [2], the attacker bypasses the authentication at the visited network's GMLC by impersonating the GMLC to the MSC/VLR. The message flow for the attack using LCS messages, also shown in Figure 8:

1. The attacker needs to know the victim's IMSI and MSC/VLR GT. The IMSI can be obtained by MAP_SRI as described above.
2. Now the attacker queries the MSC/VLR in the visited network for the accurate location information. He sends the MAP_PSL request to the MSC/VLR. The MSC/VLR has not means for authenticating this request because the LCS client authentication should have taken place already at the GMLC, which the attacker bypassed with the direct request to the MSC/VLR.
3. The MSC/VLR detects the mobile station's location with one of the various possible methods, such as the RRLP Request to the mobile. It then responds to the attacker with the MAP Provide Subscriber Location Response, which reveals the location of the victim to the attacker.

E. Experiment paradigm

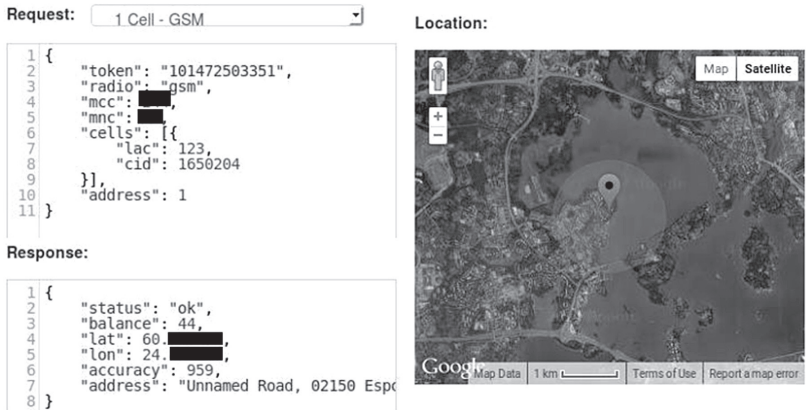
Being one of the network providers, in cooperation with an operator, we had access to the core network honeypot system. Since this honeypot was built using virtualisation of the kernel of actual core network nodes, we used it for confirming the attacks. Using such protected test environments not only avoids disruption of actual telecom services while conducting experiments, but also it helps in practical realisation of feasibility of the attacks on the real infrastructure by continuous monitoring. Our setup had one HLR, SMSC, EIR, and two MSC/VLR (of which we used one of the MSCs as part of home and the other as the visited network's VLR) which was fine-tuned to use our test SIMs. To mimic the actual attack scenarios, we kept specific ports of these elements open so that we could avoid rudimentary steps such as port scanning or topology mapping during our experiments. Core network nodes were connected to

each other using Stream Control Transmission Protocol (SCTP) and some of the SCTP ports were also open to interact with IP Internet. Since attackers use such ports an entry point to SS7, we replicated the same scenario to inject crafted packets from VLR to other elements of home network in our setup. We established SCTP client-server connection over known ports using PySCTP modules, generated SS7 traffic using the Seagull tool, and used Scapy, the packet manipulation tool, to inject false IMSIs and GTs of MSCs into the traffic of regular protocol messages exchanged between the nodes.

Following this, we monitored the packets throughout its journey using Wireshark to track and confirm the practicality of our attacks. Since our setup modelled real world scenarios, we were able to confirm the attacks presented in section 3A, 3B and the hybrid attack in 3C. For legal reasons we were not allowed to use the LCS messages (section 3D) and hence we could not confirm the feasibility of this attack. Furthermore, our test setup rejected answering ATI messages for security reasons and for the same reason we could not confirm the first attack described in section 3C.

The attackers can buy core network access from untrustworthy operators, if not portraying themselves as genuine operators by being virtual operator networks, thereby misusing the loaned infrastructure. In such cases, they can carry out attacks similar to those in our experiments. After gaining access to the core, they can use openly available tools such as SCTPscan (for port scanning), SS7Calc (for topology mapping) and Hydra (to brute force passwords) along with the tools that we used in our setup to exploit the system. Figure 9 demonstrates the mapping of authors' Cell IDs to their latitude and longitude, which is done using an online Application Programming Interfaces (API) [17]. Similarly, in many countries, the coordinates of the cell towers are public information. An attacker can either use such information if available, or use APIs such as[17] to visualise the location of the victim once he retrieves the Cell IDs.

FIGURE 9: MAPPING OF CELL ID USING THIRD PARTY API [17]



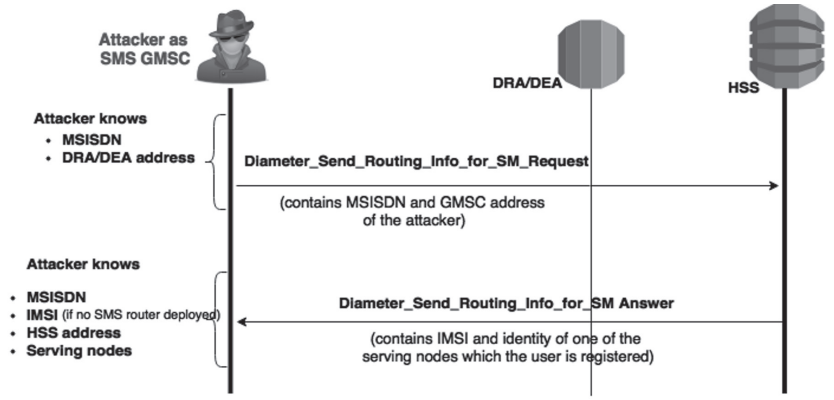
We speculate that the SS7 attackers extensively use SIGTRAN to find their entry point to the core network. Being an adjunct of the SS7 suite, SIGTRAN supports call management and application exemplars of SS7 but over IP and SCTP. SIGTRAN also facilitates adaptation of VoIP networks to the PSTN signalling. Attackers use the open interfaces of SCTP to IP dig deeper into the network.

In our understanding, the attacker would need more time to identify the open ports and map the network periphery, compared to executing the attacks themselves, which takes less time. The major costs of such attacks lie in gaining access (either by buying access or using femtocells) rather than in executing the attacks which depends only on the attackers' skillset and selection of tools. Due to these variable factors, we cannot estimate the economic feasibility of the attacks from an attacker's point of view.

4. EVOLUTION

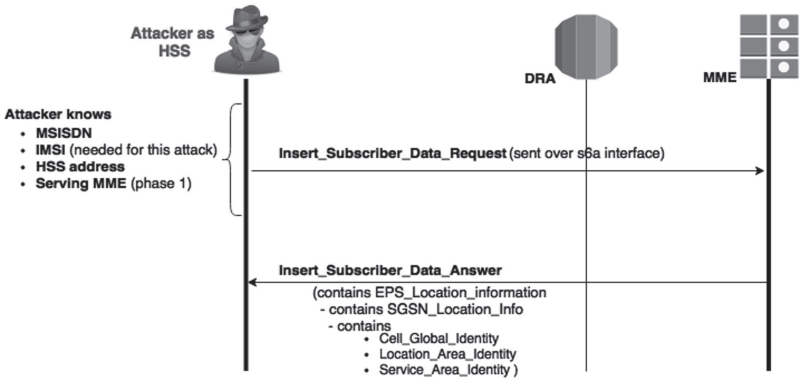
Diameter is the evolution of the SS7 and MAP protocol that is used within and between the 4G Long Term Evolution (LTE) networks. It uses the Diameter Edge Agent (DEA) that resides on the border of the network as the first contact point for messages coming over the interconnection link. In this context, Home Subscriber Server (HSS) is the evolved HLR and Mobility Management Entity (MME) can be considered to be an evolved MSC.

FIGURE 10: IMSI ACQUISITION IN LTE



Based on the existing SS7 attacks, we see the attack evolution in LTE. The potential steps are outlined in Figures 10 and 11 and can be seen as an evolution of the SS7 attack in [1]. First the attacker needs to obtain the users' IMSI and used SMS related protocol messages over the S6c interface. The attacker then starts the real part of the attack where he uses the diameter Insert Subscriber Data (ISD) command that is configured to request the location of the user over S6a interface.

FIGURE 11: LOCATION ACQUISITION IN LTE



This approach mirrors the MAP based Provide Subscriber Information approach for Diameter. LTE roaming interfaces are not yet common, and further practical testing of Diameter-based vulnerabilities are part of our ongoing research.

5. COUNTERMEASURES

One obvious solution to defend the system against the attacks discussed in section 4 is by authenticating the SS7 signalling messages. If SS7 runs over IP, then IPsec should be considered [18]. Another approach is to block or filter messages that give out location, based on the origin of the message. These may not be effective solutions as many of the signalling messages used for the attacks are part of basic communication such as calls and SMS. Filtering these messages might affect the performance of cellular services. Nevertheless, properly deployed filters can help weed out many attacks.

Firstly, the MAP_ATI command should only be used network internally within the operator network, which allows blocking of MAP_ATI requests coming over the interconnection. Secondly, both the MAP_PSI and MAP_PSL messages should be filtered to prevent bypassing of higher-level authentication. Cross-layer checks, in particular for the source address information between the Signalling Connection Control Part (SCCP) transport layer and in the MAP application layer, could help to detect spoofing of signalling messages and detection of an attack.

It is hardest to filter the MAP_SRI and MAP_SRI_SM. Even when these originate from an unknown source, they may be required for normal function of the network, such as call setup or SMS delivery. Fortunately, the attacks that exploit these messages only reveal the mobile's location at the MSC Service Area level [19].

The most promising solution for preventing the leakage of the MSC/VLR and MSRN is called SMS home routing [20], which routes the communication, via the home network without providing any location information to the sender. It requires the MAP Mobile Terminated Forward SM message to be routed always through an SMS router in the home network. Rather than revealing the IMSI of the receiving MS, the MAP_SRI response from the HLR will only contain a 15-digit MT-SMS Correlation ID [20]. This number establishes a mutual relationship between the MAP_SRI_SM and the MAP Mobile Terminated Forward SM messages without revealing the IMSI to the SMS message source. There are, however, some objections to the home routing. First, operators are used to charging more for the so-called transparent mode of SMS delivery where the SMS delivery reports are correctly returned to the sender. Second, some global operators depend on the IMSI in the current SMS delivery reports for the implementation of their billing system. Additional proposed countermeasures can be found in [21].

6. CONCLUSION

The SS7 attacks discussed in this paper make use of the lack of security in signalling protocols to breach mobile user's location privacy, whereas the Diameter attack as an evolution of SS7-based attacks, hints that the vulnerabilities are persistent in newer generations of networks. We have reviewed these attacks and explained how they work on the level of exact protocols and messages. The attacks presented in this paper have been confirmed in a real 3G/LTE network, and this confirmation and evolution of the previous knowledge to LTE networks is one of the main contributions this paper.

The initial attacks described enable an attacker with SS7 access to retrieve the IMSI and the GT of the MSC/VLR based on MSISDN, without alerting of the victim. While the MSC/VLR GT will only disclose the approximate location of the victim, it can be further narrowed down to cell area or better using the attacks presented in later part of section 3. Though Diameter seems to be an improvement over SS7 in terms of security with the use of IPsec or TLS and certificate based authentication, it is possible to port SS7 attacks to Diameter. Additionally, backward compatibility needs to be ensured between these networks, and hence downgrading attacks will remain as a persistent threat to the telecommunication industry.

Telecommunication networks are intricate systems made up of diverse circuitous subsystems, each of which comprises various different technologies. While legacy sub-systems and components are here to stay for many years, it is important to remember that the security of the whole system depends on the security of the weakest link and partner.

REFERENCES

[1] Tobias Engel, 'Locating Mobile Phones using Signaling System 7', 25th Chaos Communication Congress 25C3 (2008), http://berlin.ccc.de/~tobias/25c3-locating-mobile-phones.pdf.

[2] Tobias Engel (Sternraute), 'SS7: Locate. Track. Manipulate', 31st Chaos Communication Congress 31C3 (2014), http://berlin.ccc.de/~tobias/31c3-ss7-locate-track-manipulate.pdf.

[3] SR Labs, 'SnoopSnitch,' Security Research Lab, https://opensource.srlabs.de/projects/snoopsnitch.

[4] Karsten Nohl (SR Labs), 'Mobile self-defense' 31st Chaos Communication Congress 31C3 (2014), https://events.ccc.de/congress/2014/Fahrplan/system/attachments/2493/original/Mobile_Self_Defense-Karsten_Nohl-31C3-v1.pdf.
[5] Alexandre De Oliveira et.al. 'Worldwide attacks on SS7 network' Hackito Ergo Summit (2014), http://2014.hackitoergosum.org/slides/day3_Worldwide_attacks_on_SS7_network_P1security_Hackito_2014.pdf.
[6] Sergey Puzankov, Dmitry Kurbatov (Positive Technologies), 'How to Intercept a Conversation Held on the Other Side of the Planet', PHDays (August 2014), http://2014.phdays.com/program/tech/36930/.
[7] Australian TV Channel 9, 60 Minutes Show, 'Special Investigation: Bugged, Tracked, Hacked', (August 2015), http://www.9jumpin.com.au/show/60minutes/stories/2015/august/phone-hacking/.
[8] 3rd Generation Partnership Project (3GPP), TS 29.002, 'Mobile Application Part (MAP) specification,' Release 13, 2015, http://www.3gpp.org/DynaReport/29002.htm.
[9] Federal Communication Commission, '911 Wireless Services', 2015, https://www.fcc.gov/guides/wireless-911-services.
[10] 3rd Generation Partnership Project (3GPP), TS 44.031 'Location Services (LCS); Mobile Station (MS) - Serving Mobile Location Centre (SMLC) Radio Resource LCS Protocol (RRLP) ', Release 13, 2015, http://www.3gpp.org/DynaReport/44031.htm.
[11] Mouly, M. Pautet, 'The GSM system for mobile communications', Palaiseau, France: Cell & Sys, 1992, pp. 100-472.
[12] Shodan.io, 'SHODAN – search engine for internet connected devices', 2015, http://www.shodan.io/
[13] Learntelecom.com, 'SMS in GSM Network | Learn Telecom', 2010, http://learntelecom.com/sms-in-gsm-network/.
[14] 3rd Generation Partnership Project (3GPP), TS 23.078, 'Customised Applications for Mobile network Enhanced Logic (CAMEL)', Release 13, 2015, http://www.3gpp.org/DynaReport/23078.htm.
[15] C. Pudney, 'Liaison Statement on Restoration of R'96 Any Time Interrogation functionality', 3GPP TSG-SA WG2 meeting #22, 2002.
[16] 3rd Generation Partnership Project (3GPP), TS 23.271 'Location Services (LCS); Functional description; Stage 2', Release 13, 2015, http://www.3gpp.org/DynaReport/23271.htm.
[17] Unwired Labs, 'Unwired Labs Location API - Geolocation API and Mobile Triangulation API, Cell Tower database', Unwired Labs Location API - Geolocation & Mobile Triangulation API, 2015, http://unwiredlabs.com/api.
[18] 3rd Generation Partnership Project (3GPP), TS 33.210 '3G security; Network Domain Security (NDS); IP network layer security', Release 13, 2015, http://www.3gpp.org/DynaReport/33210.htm.
[19] Positive Technologies, 'Signaling System 7 (SS7) security report ', December 2014.
[20] 3rd Generation Partnership Project (3GPP), TR 23.840, 'Study into routing of MT-SMs via the HPLMN', Release 7, 2007, http://www.3gpp.org/DynaReport/23840.htm.
[21] GSMA permanent reference document (PRD) FS.07 'SS7 and SIGTRAN Network Security v1.0' (Currently internal to the GSMA association, but to be released to the public later.).
[22] Rao, Siddharth Prakash, et al. 'Unblocking Stolen Mobile Devices Using SS7-MAP Vulnerabilities: Exploiting the Relationship between IMEI and IMSI for EIR Access.' Trustcom/BigDataSE/ISPA, 2015 IEEE. Vol. 1. IEEE, 2015.

BIOGRAPHIES

Editors and Co-Editors

Maj **Pascal Brangetto** is a supply officer in the French Army. He graduated from the Military Administration Academy in 2006 and served as a 1st lieutenant at the 4th French Foreign Legion Batallion as a deputy administrative and supply officer. Then he went on to serve as an administrative and supply officer at the 1st Medical Battalion in Metz and was deployed for a tour of duty in Afghanistan during the ISAF operation in 2010. Before being posted as a legal researcher at NATO CCD COE in 2013, he was a staff officer in the French Joint Supply Service Directorate in Paris. Major Brangetto is a graduate from the Institut d'Etudes Politiques in Aix-en-Provence.

Lauri Lindström is a researcher at NATO CCD COE since May 2013. Prior to joining NATO CCD COE he worked at the Estonian Ministry of Foreign Affairs (2007-2012) as the Director General of Policy Planning and held various positions at the Ministry of Defence (1995-2007) dealing mainly with issues related to international cooperation, Estonia's accession to NATO, defence planning and security policy. Lauri Lindström holds a PhD from the Tallinn University, Estonia.

Maj **Nikolaos Pissanidis** is a Greek Army IT officer with more than 20 years of professional experience in the field of IT and IT security. Before his current assignment as a researcher at NATO CCD COE's Technology Branch, he worked in several different national and international management, leadership and expert positions focusing on information technology, software development, cyber security and web penetration testing. Besides a diploma from the Hellenic Army Academy, Niko holds a master's degree in New Technologies in Informatics and Telecommunications from the Department of Informatics and Telecommunications of National and Kapodistrian University of Athens.

Henry Rõigas is a researcher in the Law and Policy Branch at the NATO CCD COE. His research in the Centre focuses mainly on the political aspects of international cyber security. Henry was the co-editor of the book *International Cyber Norms: Legal, Policy & Industry Perspectives* and Project Manager of the book *Cyber War in Perspective: Russian Aggression against Ukraine*. He also manages the Centre's INCYDER (International Cyber Developments Review) database and co-organises CyCon. Henry holds a Master's degree in International Relations from the University of Tartu.

Matthijs Veenendaal has been working for the Netherlands Ministry of Defence since 2006 in various policy positions. He is currently stationed as a researcher at the Strategy Branch of the NATO CCD COE in Estonia. He has been closely involved in the development of cyber defence policy of the ministry of Defence and the principle author of the first Defence Strategy for operating in cyberspace (2012). He was also closely involved in the development of the two National Cyber Security Strategies of the Netherlands. Matthijs studied contemporary history at the university of Leiden and political science at the University of Texas, Austin.

Teemu Väisänen is working as a researcher for NATO CCD COE in Tallinn, Estonia, and studying for PhD at the University of Oulu, Finland. He has more than 11 years of experience in information security research and development in projects consisting of different topics of information and cyber security. He has been working as a Researcher for Finnish Defence Forces and for VTT Technical Research Centre of Finland (VTT), as a voluntary Safety Expert of VTT in Safer Internet Day (SID) project led by Finnish Communications Regulatory Authority, and as one of VTT's voluntary Mediataitokummi in Media Literature School project led by the National Audiovisual Institute (of Finland).

Authors

Anthony Craig is a PhD student at Cardiff University, and also a member of the Research School on Peace and Conflict at the PRIO in Oslo. He holds an MA (Hons.) in History and Politics, and an MRes in International Relations, from the University of Glasgow. His main interests lie in the causes and consequences of military build-ups and arms races in international politics. Anthony's current research involves collecting data on nation states' cyber capabilities as he seeks to relate these ideas to the cyber domain. His recent publications include a paper on states' reactions to cyber threats in terms of the acquisition of encryption technologies.

Jelle van Haaster is currently writing his multidisciplinary PhD thesis on the future utility of military cyber operations during conflicts at the Faculty of Military Sciences Breda and University of Amsterdam. Apart from being an officer in the Royal Netherlands Army, he has a background in international law (LL.M.) and software development (big data analytics, serious games). He has received various awards for his accomplishments in software development and academia.

Kim Hartmann has been employed at the Institute of Electronics, Signal Processing and Communication at Otto von Guericke University, Magdeburg, Germany, since 2011. Kim Hartmann conducts applied research in secure network design principles, risk analysis and assessment of networks, network components, and protocols, and provides policy recommendations based on sound technical understanding of the realities of cyber threat and network vulnerabilities. She is a regular contributor to research projects and conferences on cyber and network security. Her academic work specializes in computer security and mathematical modelling, protocol security analysis, computer security risk assessment, and risk analysis of critical network infrastructures. Kim Hartmann studied Computer Science and Mathematics at the Royal Institute of Technology, Stockholm, Sweden; and Otto von Guericke University, Magdeburg, Germany. She is registered as a 'Horizon 2020' (EU Research and Innovation programme) expert in cyber security, communications and network security, and human-computer interaction (HCI) and has been active as a freelancer in the computer security consulting field since 2010.

Jason Healey is a senior research scholar at Columbia University's School for International and Public Affairs specialising in cyber conflict, competition and cooperation. Prior to this, he was the founding director of the Cyber Statecraft Initiative of the Atlantic Council where

he remains a Senior Fellow. He has authored dozens of published articles and is the editor of the first history of conflict in cyberspace, *A Fierce Domain: Cyber Conflict, 1986 to 2012*. During his time in the White House, he was a director for cyber policy, coordinating efforts to secure US cyberspace and critical infrastructure. At Goldman Sachs, he created their first cyber incident response team and later oversaw the bank's crisis management and business continuity in Asia. He started his career as a US Air Force intelligence officer with jobs at the Pentagon and National Security Agency and is president of the Cyber Conflict Studies Association.

Trey Herr is a fellow with the Belfer Center's Cyber Security Project at the Harvard Kennedy School. His work focuses on trends in state developed malicious software, the structure of criminal markets for malware components, and the proliferation of malware. Trey is also a non-resident fellow with New America's Cybersecurity Initiative and an adjunct researcher with the Institute for Defense Analyses. He holds a PhD and MA in Political Science from George Washington University and a BS in Theatre and Political Science from Northwestern University. He is co-editor of *Cyber Insecurity*, a multi-author volume on cyber security policy forthcoming in Fall, 2016.

Drew Herrick is a Political Science PhD candidate specialising in international relations and research methods at George Washington University. He is also a Nonresident National Cybersecurity Fellow at New America. His research primarily focuses on the intersection of international security and technology, especially counter-norm activity, offensive social media operations, and the military value of cyber capabilities.

Robert Koch is a research assistant in the Department of Computer Science at Universität der Bundeswehr München and a member of the university's Research Center for Cyber Defence (CODE). His research interests include network and system security with an emphasis on intrusion and extrusion detection. Koch received a PhD in informatics from Universität der Bundeswehr München.

Sean Lawson is associate professor in the Department of Communication at the University of Utah, USA.

Kubo Mačák is a lecturer in Law at the University of Exeter, UK. He holds the degrees of DPhil, MPhil, and MJur from the University of Oxford, and an undergraduate degree in law (Magister) from Charles University in Prague. In 2012, he was awarded the Diploma of the Hague Academy of International Law. Kubo has held research positions at the Universities of Bonn (Germany), Haifa (Israel), and Wuhan (China). He has worked at the United Nations ad hoc tribunals for the former Yugoslavia and Rwanda. His research interests span the law of cyber security, international humanitarian law, and general international law.

Mirco Marchetti is a researcher of the Inter-department Research Center on Security (CRIS) at the University of Modena and Reggio Emilia, Italy. He received the Laurea degree and the PhD in Computer Engineering from the University of Modena with projects on large scale systems for information security. He is an expert in cooperative network intrusion detection

and prevention, fault tolerant distributed systems, high performance systems for information security management and cloud security. He is a teacher in the Master in Cyberdefense organized by the Armed Forces Institute of Telecommunications and the University of Modena and Reggio Emilia.

Markus Maybaum is a German Air Force officer with more than 20 years of professional experience in the field of IT and IT security. He worked in several different national and international management, leadership and expert positions focusing on information technology, software engineering, cyber security and arms control. Currently, Markus is heading the international relations office for cyber security at the German military IT Service Centre and he is working for Fraunhofer FKIE's Cyber Analysis & Defense department as a researcher in the field of cyber arms control and trusted architectures. Before his current assignments, Markus was assigned to NATO CCD COE's Technology Branch as a researcher where he used to manage CyCon's Technical Track for a couple of years. He was also the course director of the Centre's Botnet Mitigation Training and the Malware & Exploitation course. Markus is still actively contributing to the mission of the Centre as a NATO CCD COE ambassador. Besides a diploma in business administration from the German Air Force College, Markus holds a master's degree in informatics from the German Open University of Hagen specialising in IT security and he is currently pursuing a PhD in information technology with a focus on technical aspects of arms control in cyberspace at Fraunhofer FKIE and the University of Bonn, Germany.

Paolo Palmieri is a lecturer in the Department of Computing and Informatics and a member of the Cyber Security Research Group at Bournemouth University (UK). He holds a PhD in cryptography from the Université Catholique de Louvain (Belgium), and has been a Visiting Researcher at the Università di Bologna (Italy) and a Post-Doctoral Researcher at Delft University of Technology (The Netherlands). His research interests space from cryptographic protocols to privacy and anonymity, and include privacy enhancing technologies, secure computation, location privacy, and the security of smart cities and the Internet of Things.

Siddharth Prakash Rao is a research student from "Secure Systems" research group of Aalto University, Finland. Prior to that he has worked at security research teams of Bell Labs - Nokia, Helsinki Institute of Information Technology (HIIT) and Fidelity Management and Research. As an Erasmus Mundus student he holds double master's degrees in security and cryptography from Aalto University, Finland and University of Tartu, Estonia respectively. His research interests include security and privacy of network protocols, cellular core network vulnerabilities, pedagogical study of emerging threats to Internet privacy problems.

Ragnhild Endresen Siedler is a researcher at Norwegian Defence Research Establishment, Analysis Division. Her research area is military operations and cyber defence. She holds a master's degree in political science from the University of Oslo. Before her current position, she served as an officer in Royal Norwegian Air Force and also holds a bachelor's degree in military studies from Royal Norwegian Air Force Academy. Siedler is a Norwegian representative to the Multinational Capability Development Campaign and undertook project lead responsibilities in the 2013-2014 campaign.

Steve Sin is a lead investigator and senior researcher in the Unconventional Weapons and Technology Division (UWT) of the National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland. His current research focuses on illicit trafficking of radiological and nuclear material; non-state actor cyber capabilities; and non-state actor technology diffusion, transfer, and adoption. Prior to his time at START, He served in the U.S. Army as a Counterintelligence Officer specialising in counter-terrorism and political-military affairs in the Asia-Pacific Theater of Operations. He is also a PhD candidate in political science at the State University of New York at Albany, and his dissertation examines the relationship between interstate crisis initiation and distraction.

Martin Strohmeier is a final-year DPhil student and teaching assistant in the Department of computer Science at the University of Oxford. His main research interests are currently in the area of network security, including wireless sensor networks and critical infrastructure protection. During his time at Oxford, he has extensively analysed the security and privacy of wireless aviation technologies of this generation and the next. His work predominantly focuses on developing cyber-physical approaches which can improve the security of air traffic control quickly and efficiently. He has received several best paper awards from both the aviation and computer security community and is a co-founder of the aviation research network OpenSky. Before coming to Oxford in 2012, he received his MSc degree from TU Kaiserslautern, Germany and joined Lancaster University's InfoLab21 and Lufthansa AG as a visiting researcher.

Lior Tabansky is a scholar of cyber power at Tel Aviv University's Blavatnik Interdisciplinary Cyber Research Center (TAU ICRC) and the Director of Cyber Strategy with the consultancy firm CSG. Mr. Tabansky's doctoral dissertation in Political Science develops Military Innovation framework towards (inter)national cybersecurity integration with Strategic Studies. *Cybersecurity in Israel*, his book co-authored with Prof. Isaac Ben-Israel, provides the first systematic account of Israeli cyber power. His commitment to rigorous academic research uniquely builds upon practice, from his service in the Israeli Air Force, through a corporate IT career, policy efforts at the INSS think-tank and high-level cybersecurity consulting.