2012 4th International Conference on Cyber Conflict

PROCEEDINGS

C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)

CCDCOE
NATO Cooperative Cyber Defence Centre of Excellence Tallinn Estonia

IEEE Advancing Technology for Humanity

IEEE COMMUNICATIONS SOCIETY

5-8 JUNE, 2012 TALLINN, ESTONIA

2012 4TH INTERNATIONAL CONFERENCE ON CYBER CONFLICT (CYCON 2012)

IEEE Catalog Number:	CFP1226N-PRT
ISBN 13 (print):	978-9949-9040-8-2
ISBN 13 (pdf):	978-9949-9040-9-9
ISBN 13 (epub):	978-9949-9211-0-2

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

4TH INTERNATIONAL CONFERENCE ON CYBER CONFLICT SPONSORS

The NATO CCD COE and the conference organisers want to thank the following sponsors for their support of this year's conference:

ABOUT THE NATO CCD COE

The NATO Cooperative Cyber Defence Centre of Excellence (NATO CCD COE) is an international military organisation accredited in 2008 by NATO's North Atlantic Council as a "Centre of Excellence". Located in Tallinn, Estonia, the Centre is currently supported by Estonia, Germany, Hungary, Italy, Latvia, Lithuania, the Netherlands, Poland, Slovakia, Spain, and the USA as Sponsoring Nations. The Centre is not part of NATO's command or force structure, nor is it funded by NATO. However, it is part of a wider framework supporting NATO Command Arrangements.

The NATO CCD COE's mission is to enhance capability, cooperation and information-sharing between NATO, NATO member States and NATO's partner countries in the area of cyber defence by virtue of research, education and consultation. The Centre has taken a NATO-orientated, interdisciplinary approach to its key activities, including academic research on selected topics relevant to the cyber domain from legal, policy, strategic, doctrinal and/or technical perspectives, providing education and training, organising conferences, workshops and cyber defence exercises and offering consultations upon request.

For more information on the NATO CCD COE, please visit the Centre's website at http://www.ccdcoe.org.

For information on Centres of Excellence, visit NATO's website "Centres of Excellence" at http://www.nato.int/cps/en/natolive/topics_68372.htm.

FOREWORD

Protecting critical information assets, enabling safe communications, and conducting effective military operations in cyberspace have become critical national security priorities for many nations. They are priorities driven not only by dramatic advances in cyber attack tools and technology, but also by the dangerous reality that such capabilities are in the hands of governments, criminal organizations, terrorist groups, and individual hackers. Once isolated and relatively ineffective, cyber attacks today are highly effective and well-coordinated operations that often consist of sequenced tit-for-tat attacks and counter-attacks. Such attacks have the theoretical potential to paralize national economies, cripple governmental functions, and endanger the physical well-being of populations. The inevitable result of these developments could well be local or even global cyber conflagrations, with consequences rising to the level of those occurring during traditional physical military confrontations.

Academic institutions, industrial fora, and governmental organizations have been conducting research, performing case studies, and modeling cyber operations for a number of years. In this active cyber security "ecosystem", the International Conference on Cyber Conflict (CyCon) conducted annually in Tallinn by the NATO Cooperative Cyber Defence Centre of Excellence has emerged as especially influential, for they offer a unique perspective. This distinctiveness is marked by an innovative synergistic approach to the conceptual framework, architectures, processes and systems of cyber security and conflict. It holistically examines strategic and policy matters, social and economic concerns, law, computer science and Information technologies, military doctrine, and human behavioral modeling with respect to cyber space. No less important is the open and interactive forum it offers to world-class researchers, legal/policy/military experts, and IT practitioners.

The Proceedings of CyCon 2012, conducted with the technical support of the IEEE Communications Society, are collected in this volume. The twenty-nine papers were selected by the conference Programme Committee following a rigorous peer review process conducted by distinguished experts. The works are sprinkled across the legal, policy, strategic, and technical spectra of cyber conflict; they include sophisticated analyses of topics like offensive and defensive cyber activities, the concept of the cyber space, its legal and technical boundaries, and the fundamental notions of cyber attacks, cyber attackers, cyber conflict, and cyber warfare. We hope readers will agree that they comprise erudite and useful examinations of the key questions with which scholars and practitioners continue to struggle.

This volume is arranged into six chapters. The first, *Cyberspace – The Role of States in the Global Structure*, includes articles that consider the role of governments in cyber space policy-making. The second chapter, *Cyber Policy & Strategic Options* builds on the first with papers that turn to the policies and strategies that States are adopting on such matters as cyber space protection, militarization of the cyber space and cyber warfare. Articles in the third chapter, *Cyber Conflict – Theory & Principles*, examine theoretical issues and the historical and operational context in which they are at play. The fourth chapter, *Cyber Conflict – Actors*, is a collection of papers that identifies the various human and software actors involved in cyber

conflict and examines their roles and objectives, as well as the ramifications of their activities. Chapter Five, *"Cyber-Attacks" – Trends, Methods & Legal Classification*, deals with the activities of those cyber actors. It includes articles on computer network attack, methods of commanding and controlling cyber attacks, and cyber attacks in the context of international law. The volume concludes with a chapter, *Cyber Defence – Methods & Tools*, providing a series of technical pieces on cyber security information sensing, decision support and tools supporting the tasks of cyber defense.

We would like to thank the distinguished members of the CyCon 2012 Programme Committee for their tireless work in identifying papers for presentation at the conference and publication in this book. Most importantly, though, we are delighted to congratulate this volume's editors – Dr Katharina Ziolkowski, CPT Christian Czosseck, and Dr Rain Ottis of the NATO Cooperative Cyber Defence Centre of Excellence. Without their technical expertise, professional attitude, personal dedication, and boundless enthusiasm this impressive work would not have been possible.

The CyCon 2012
Programme Committee Co-Chairs

Dr Gabriel Jakobson
Chief Scientist, Altusys Corp
Brookline, MA

Professor Michael N. Schmitt
United States Naval War College
Newport, Rhode Island

TABLE OF CONTENTS

INTRODUCTION

For the fourth year in a row, the NATO Cooperative Cyber Defence Centre of Excellence (NATO CCD COE) invited experts from government, academia and industry to Tallinn to discuss recent trends in cyber defence. The *4th International Conference on Cyber Conflict* (CyCon 2012) brought together national security thinkers, strategists, political scientists, policy-makers, lawyers and technology experts interested in cyber defence, and served as a hub for knowledge and networking on an international level.

CyCon 2012 focused on "Military and Paramilitary Activities in Cyberspace". Reflecting the interdisciplinary approach of NATO CCD COE, this topic was explored from strategic, conceptual, political, legal and technical perspectives within two parallel tracks, the *Law & Policy Track*, chaired by Dr *Katharina Ziolkowski* (NATO CCD COE) and the *Technical Track*, chaired by Dr *Rain Ottis* (NATO CCD COE). Additionally, a *Strategy Breakout Session*, moderated by Dr *Kenneth Geers* (former member of NATO CCD COE), and multiple pre-conference workshops (organised by the Cyber Security Forum Initiative and NORMAN), completed the comprehensive programme offered.

The *Law & Policy Track* addressed the principles of territoriality, national sovereignty and neutrality in cyberspace, and the impact of the global access to cyberspace upon diplomatic relations, human rights and democratic movements or uprisings. Also, the notion of an "armed attack" with regard to cyberspace was addressed from legal and political viewpoints, as well as the topic of State responsibility, the challenge of political attribution of malicious activities and the possibilities of deterrence or compellence in cyberspace. Importantly, the *Law & Policy Track* focused on the applicability of the Law of Armed Conflict to cyberspace, addressing topics of military targeting such as *inter alia*, combatancy and direct participation in hostilities in cyberspace.

The *Technical Track* addressed the technological side of cyber conflict, cyber attack and cyber weapons. This track explored case studies of some of the most influential cyber conflicts, attacks and weapons from recent times as well as hypothetical scenarios of (cyber) conflicts. Furthermore, the *Technical Track* discussed technical challenges and solutions for attribution, one of the major challenges in cyber security, from a technological point of view, as well as tools, methods and protocols of IT security, information sharing and cyber espionage.

The *Strategy Breakout Session* explored the topics of cyber conflict and cyber warfare from a conceptual perspective. This session explored ways to carry over traditional military concepts such as manoeuvre warfare to operations in cyberspace. In addition, the session included conceptual approaches for describing cyberspace as a modern field of conflict, as well as overviews of various national viewpoints on international cyber security.

The *Joint Sessions*, addressing all participants of the conference, addressed topics such as the future of cyber conflicts, the ethical aspects of cyber warfare and the effects of militarisation of cyberspace.

The editors of these Conference Proceedings, i.e. the Track Chairs and the Publication Chair *Christian Czosseck* (NATO CCD COE), have undertaken the challenge to not structure the publication in accordance with the disciplines reflected in the conference's tracks. Instead, we have rearranged the articles into categories of specific themes, which are explored from the perspectives of the different disciplines over the course of the conference, as this reflects best the interdisciplinary approach that NATO CCD COE and CyCon 2012 aspire to.

The editors would like to thank the Co-Chairs and distinguished members of the Programme Committee for their efforts in reviewing, discussing and selecting the papers submitted pursuant to the "Call for Papers", and also for the peer review of the papers submitted by invited authors, guaranteeing the academic quality of the selected papers.

Programme Committee Co-Chairs were (in alphabetic order):

- Dr *Gabriel Jakobson*, Chief Scientist, Altusys Corporation
- Prof Dr *Michael N. Schmitt*, Chairman of the International Law Department, U.S. Naval War College

Members of the Programme Committee were (in alphabetic order):

- Dr *Iosif I. Androulidakis*, University of Ioannina, Greece
- Prof. *Marta Beltrán*, Rey Juan Carlos University, ESP
- Air Cdre (ret.) Dr *William Boothby*, UK
- Dr *Catharina Candolin*, Defence Forces, Finland
- Prof *Tom Chen*, Swansea University, UK
- Mr *Christian Czosseck*, NATO CCD COE
- Prof *Dorothy Denning*, Naval Postgraduate School, USA
- Prof Dr *Chris C. Demchak*, US Naval War College, USA
- Dr *Myriam Dunn Cavelty*, Center for Security Studies, Swiss Federal Institute of Technology, Switzerland
- Dr *Kenneth Geers*, Naval Criminal Investigative Service, USA
- Prof Dr *Robin Geiss*, University of Potsdam, Germany
- Prof Dr *Terry D. Gill*, University of Amsterdam, The Netherlands
- Prof Dr *Michael Grimalia*, Air Force Institute of Technology, USA
- Mr *Jason Healey*, Atlantic Council, USA
- Prof Dr *Wolff Heintschel von Heinegg*, Europe-University Viadrina, Germany
- Prof *Eric Talbot Jensen*, Brigham Young University, USA
- Dr *Marieke Klaver*, Netherlands Organisation for Applied Scientific Research TNO, The Netherlands
- Mr *Latif Ladid*, IPv6 Forum, Luxembourg
- Dr *Pavel Laskov*, University of Tübingen, Germany
- Dr *Corrado Leita*, Symantec, France
- Assoc. Prof *Samuel Liles*, National Defense University, USA
- *Eric Luiijf*, M.Sc (Eng) Delft, Netherlands Organisation for Applied Scientific Research TNO, The Netherlands
- Dr *Jose Nazario*, Arbor Networks, USA

- Mr *Lars Nicander*, National Defence College, Sweden
- Dr *Rain Ottis*, NATO CCD COE
- Assoc. Prof Dr *Julie Ryan*, George Washington University, USA
- Prof Dr *Noel Sharkey*, University of Sheffield, UK
- Mrs *Heli Tiirmaa-Klaar*, European Union, European External Action Service, Belgium
- Prof Dr *Enn Tõugu*, NATO CCD COE and Tallinn University of Technology, Estonia
- Dr *Risto Vaarandi*, NATO CCD COE and SEB Bank, Estonia
- Dr *Jozef Vyskoč*, VaF Rovinka and Comenius University Bratislava, Slovak Republic
- Assoc. Prof *Sean Watts*, Creighton University, USA
- Prof *Stefano Zanero*, Politecnico di Milano, Italy
- Dr *Katharina Ziolkowski*, NATO CCD COE

Special gratitude is due to the Institute of Electrical and Electronics Engineers (IEEE) Communications Society, the world's largest professional association dedicated to advancing technological innovation and excellence for the benefit of humanity. The IEEE's Communications Society served as technical co-sponsor of CyCon 2012 and of these Conference Proceedings, ensuring the academic quality of the papers and supporting their electronic publication and distribution.

Last but not least, we would also like to thank all authors of the papers collated in this publication for their superb submissions and friendly cooperation during the course of the publication process.

Dr *Katharina Ziolkowski*, *Christian Czosseck* & Dr *Rain Ottis*
NATO Cooperative Cyber Defence Centre of Excellence

Tallinn, Estonia
June 2012

Chapter 1
Cyberspace – The Role of States in the Global Structure

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Legal Implications of Territorial Sovereignty in Cyberspace

Wolff Heintschel von Heinegg
Faculty of Law
Europa-Universität
Frankfurt (Oder), Germany
heinegg@europa-uni.de

Abstract: The principle of territorial sovereignty applies to cyberspace and it protects the cyber infrastructure located within a State's territory. States are prohibited to interfere with the cyber infrastructure located in the territory of another State. This certainly holds true if the conduct is attributable and if it inflicts (severe) damage on the integrity or functionality of foreign cyber infrastructure. Moreover, States have the obligation not to allow knowingly their territory to be used for acts that violate the territorial sovereignty of another State. It is, however, unsettled whether there is a rebuttable presumption of knowledge if the cyber attacks were launched from the government cyber infrastructure of the State of origin.

States have a right to exercise their territorial jurisdiction over cyber activities within their territories. However, the characteristics of cyberspace and the necessity to preserve the functionality of the Internet call for consensual limitations of an exercise of territorial jurisdiction. The U.S. International Strategy for Cyberspace has the potential of guiding governments in order to either progressively develop international law or to specify existing norms of international law.

The attribution of cyber attacks to a given State continues to be a challenging problem. Nevertheless, States should continue to improve their capabilities in the area of cyber forensics. The U.S. Department of Defense Cyberspace Policy Report is to be considered a valuable contribution to that effect.

Keywords: *territorial sovereignty, exercise of jurisdiction, cyber infrastructure, obligations of States in cyberspace*

1. INTRODUCTION

The question whether traditional rules and principles of international law apply to conduct in cyberspace is far from new. Still, at least in Europe governments do not seem to have shown a specific interest in a clarification of the applicable norms of international law before the cyber attacks on Estonia in 2007 and on Georgia in 2008 although the discussion in the United

States of America had been underway since the end of the 20th century. Of course, it is a positive development that the issue of the applicability of (customary) international law to cyberspace has gained the attention it deserves. Less positive is the mystification of cyberspace as a 'fifth dimension' or as a 'fifth domain' that according to some is considered so novel that it eludes the traditional rules and principles of international law. Such an exaggeration of cyberspace is neither justified nor necessary and it therefore does not justify the various calls for 'new norms of international law' specifically designed for cyberspace. International law as it currently stands need not capitulate in view of the challenges brought about by cyberspace and the technology it is based upon. States seem to agree that customary international law is, in principle, applicable to cyberspace although there may be a need for a consensual adaptation to the specific characteristics of cyberspace.

The present paper will, for obvious reasons, not address the entire spectrum of customary international law that may have an impact on State conduct in cyberspace. Rather, it will explore whether and to what extent the rights and duties derived from the principle of territorial sovereignty do apply to cyberspace. It will be shown that the principle of territorial sovereignty applies to certain components of cyberspace and that the specific rights and obligations flowing from that principle have not become obsolete for the mere fact that cyberspace is characterized as a fifth dimension or as part of the global commons.

2. GENERAL CHARACTERISTICS OF TERRITORIAL SOVEREIGNTY

Irrespective of the various theories on the legal function of territory[1] there is widespread agreement that according to the principle of territorial sovereignty a State exercises full and exclusive authority over its territory.[2] Max Huber, in the Palmas Island Arbitration award, has affirmed this general principle as follows: "Sovereignty in the relations between States signifies independence. Independence in regard to a portion of the globe is the right to exercise therein, to the exclusivity of any other States, the functions of a State".[3] According to the International Court of Justice "[b]etween independent States, respect for territorial sovereignty is an essential foundation of international relations".[4] Territorial sovereignty (or: 'full and exclusive authority') therefore implies that, subject to applicable customary or conventional rules of international law, the respective State alone is entitled to exercise jurisdiction, especially by subjecting objects and persons within its territory to domestic legislation and to enforce these rules. Moreover, the State is entitled to control access to and egress from its territory. The latter right seems to also apply to all forms of communication. Territorial sovereignty protects a State against any form of interference by other States. While such interference may imply the use of force, that aspect is not dealt with here.

It must borne in mind that territorial sovereignty does not merely afford protection to States but it also imposes obligations on States, especially the "obligation to protect within the territory

1 For a discussion of the various theories on the legal function of territory see Santiago Torres Bernárdez, 'Territorial Sovereignty', in Encyclopedia of Public International Law Vol. IV, p. 823 at p. 824 *et seq*. (ed. by R. Bernhardt, Amsterdam et al. 2000).

2 See, *inter alia*, The *Lotus*, PCIJ Ser. A, No. 10, at p. 18 *et seq*. (1927); *Free Zones of Upper Savoy and Gex Case*, PCIJ Ser. A/B, No. 46, p. 166 *et seq*. (1932).

3 2 RIAA p. 829 at p. 838.

4 ICJ, *The Korfu Channel Case (Merits)*, ICJ Rep., 1, at p. 35 (1949).

the rights of other States, in particular their right to integrity and inviolability in peace and in war, together with the rights which each State may claim for its nationals in foreign territory."[5]

3. TERRITORIAL SOVEREIGNTY AND CYBERSPACE

'Cyberspace' has been defined as "a global domain within the information environment consisting of the interdependent network of information technology infrastructures, including the Internet, telecommunications networks, computer systems, and embedded processors and controllers".[6] There is a widely-held view that it "is not a physical place – it defies measurement in any physical dimension or time space continuum. It is a shorthand term that refers to the environment created by the confluence of cooperative networks of computers, information systems, and telecommunication infrastructures commonly referred to as the World Wide Web."[7] It is true that cyberspace is characterized by anonymity and ubiquity.[8] Therefore it seems logical to assimilate it to the high seas, international airspace and outer space[9], i.e., to consider it a 'global common' or legally a *res communis omnium*.[10] However, these characterizations merely justify the obvious conclusion that cyberspace in its entirety is not subject to the sovereignty of a single State or of a group of States. In view of its characteristics it is immune from appropriation.

Despite of the correct classification of 'cyberspace as such' as a *res communis omnium* State practice gives sufficient evidence that cyberspace, or rather: components thereof, is not immune from sovereignty and from the exercise of jurisdiction. On the one hand, States have exercised, and will continue to exercise, their criminal jurisdiction vis-à-vis cyber crimes[11] and they continue to regulate activities in cyberspace. On the other hand, it is important to bear in mind that "cyberspace requires a physical architecture to exist".[12] The respective equipment is usually located within the territory of a State. It is owned by the government or by corporations. It is

5 Max Huber in the *Palmas Arbitration, supra* note 3, at p. 839. In his Separate Opinion in the *Korfu Channel Case* Judge Alvarez stated: "By sovereignty, we understand the whole body of rights and attributes which a State possesses in its territory, to the exclusion of all other States, and also in its relations with other States. Sovereignty confers rights upon States and imposes obligations upon them", ICJ Rep., p. 43 (1949).

6 Joint Chiefs of Staff, Joint Pub. 1-02, Dept. of Defense Dictionary of Military and Associated Terms, at 41 (12 April 2001). See also the definition by Arie J. Schaap, 'Cyber Warfare Operations: Development and Use under International Law', 64 AFLR, 121-173, at 126 (2009), who defines 'cyberspace' as a "domain characterized by the use of [computers and other electronic devices] to store, modify, and exchange data via networked systems and associated physical infrastructures".

7 Thomas Wingfield, *The Law of Information Conflict: National Security Law in Cyberspace*, at 17 (Aegis Research Corp. 2000).

8 It has been rightly stated that "global digital networks have the features they do – of placelessness, anonymity, and ubiquity – because of politics, not in spite of them". See Geoffrey L Herrera, Cyberspace and Sovereignty: Thoughts on Physical Space and Digital Space, at 12 (2006), available at http://www.allacademic.com/meta/p98069_index.html.

9 For an analysis to that effect see Patrick W. Franzese, 'Sovereignty in Cyberspace: Can It Exist?', 64 AFLR 1-42, at 18 *et seq*. (2009).

10 U.S Department of Defense, *Strategy for Operating in Cyberspace* (available at http://www.defense.gov/news/d20110714cyber.pdf): "DoD will treat cyberspace as an operational domain to organize, train, and equip so that DoD can take full advantage of cyberspace's potential." See also U.S. Department of Defense, *The Strategy for Homeland Defense and Civil Support*, at 12 (2005:) "The global commons consist of international waters and airspace, space, and cyberspace."

11 It suffices to refer to the Council of Europe Convention on Cybercrime of 23 November 2001, E.T.S. No.185.

12 Franzese, supra note 9, at 33.

connected to the national electric grid.[13] The integration of physical components, i.e., of cyber infrastructure located within a State's territory, into the 'global domain' of cyberspace cannot be interpreted as a waiver of the exercise of territorial sovereignty. In view of the genuine architecture of cyberspace it may be difficult to exercise sovereignty. Still, the technological and technical problems involved do not prevent a State from exercising its sovereignty, especially its criminal jurisdiction, to the cyber infrastructure located in areas covered by its territorial sovereignty.

States have continuously emphasized their right to exercise control over the cyber infrastructure located in their respective territory, to exercise their jurisdiction over cyber activities on their territory, and to protect their cyber infrastructure against any trans-border interference by other States or by individuals.[14]

It needs to be emphasized that the applicability of the principle of sovereignty to the said components of, and activities in, cyberspace is not barred by the innovative and novel character of the underlying technology. This holds true for the majority of rules and principles of customary international law that do apply to cyberspace and to cyber activities. The U.S. President, in the 2011 International Strategy for Cyberspace, has clearly stated that the "development of norms for state conduct in cyberspace does not require a reinvention of customary international law, nor does it render existing international norms obsolete. Long-standing international norms guiding state behavior – in times of peace and conflict – also apply in cyberspace."[15]

This does not necessarily mean that the said rules and principles are applicable to cyberspace in their traditional interpretation. In view of the novel character of cyberspace and in view of the vulnerability of cyber infrastructure and cyber components there is a noticeable uncertainty amongst governments and legal scholars as to whether the traditional rules and principles of customary international law are sufficiently apt to provide the desired answers to some worrying questions. It is, therefore, of utmost importance that States not only agree on the principal application of customary international law to cyberspace but also on a common interpretation that takes into due consideration the "unique attributes of networked technology".[16] Hence it is necessary that governments "continue to work internationally to forge consensus regarding how norms of behavior apply to cyberspace".[17]

13 See Joshua E. Kastenberg, 'Non-Intervention and Neutrality in Cyberspace: An Emerging Principle in the National Practice of International Law', 64 AFLR, 43-64, at 64 (2009).

14 See the *Strategy for Operating in Cyberspace, supra* note 10. See further U.S. Department of Defense, *Cyberspace Policy Report - A Report to Congress Pursuant to the National Defense Authorization Act for Fiscal Year 2011*, Section 934, at 4 *et seq*. (November 2011), available at http://www.defense.gov/home/features/2011/0411_cyberstrategy/docs/NDAA%20Section%20934%20Report_For%20webpage.pdf.; U.S. President, *International Strategy for Cyberspace*, at 12 *et seq*. (May 2011).

15 *Ibid*., at 9.

16 *Ibid*.: "Nonetheless, unique attributes of networked technology require additional work to clarify how these norms apply and what additional understandings might be necessary to supplement them."

17 *Ibid*. See also the Cyberspace Policy Report, *supra* note 14, at 7: "The United States is actively engaged in the continuing development of norms of responsible state behavior in cyberspace, making clear that as a matter of U.S. policy, long-standing international norms guiding state behavior also apply equally in cyberspace. Among these, applying the tenets of the law of armed conflict are critical to this vision, although cyberspace's unique aspects may require clarifications in certain areas." At p. 9 the Report emphasizes that the "law of armed conflict and customary international law, however, provide a strong basis to apply such norms to cyberspace governing responsible state behavior."

4. SCOPE OF TERRITORIAL SOVEREIGNTY IN CYBERSPACE

The basic applicability of the principle of territorial sovereignty to cyberspace entails that the cyber infrastructure located on the land territory, in the internal waters, in the territorial sea, and, where applicable, in the archipelagic waters, or in the national airspace is covered by the respective State's territorial sovereignty.[18] Hence, in principle, the State is entitled to exercise control over that cyber infrastructure and over cyber activities in those areas. It may not be left out of consideration, however, that the exercise of sovereignty may be restricted by customary or conventional rules of international law, such as the immunity of diplomatic correspondence[19] or the rights of innocent passage, transit passage, and archipelagic sea lanes passage.[20]

A. Ratione loci

The first consequence of the above findings is that the cyber infrastructure located in areas covered by the territorial sovereignty is protected against interference by other States. This protection is not limited to activities amounting to an unjustified use of force, to an armed attack or to a prohibited intervention.[21] Rather, any activity attributable to another State, e.g. because it constitutes an exercise of that State's jurisdiction, is to be considered a violation of the sovereignty of the territorial State.[22] This also holds true if the attributable conduct has negative impacts on the integrity or functionality of the cyber infrastructure. It is important to note that not every State conduct that impacts on the cyber infrastructure of another State necessarily constitutes a violation of the principle of territorial sovereignty. If the act of interference results in inflicting material damage to the cyber infrastructure located in another State, there seems to be a sufficient consensus that such an act constitutes a violation of the territorial sovereignty of the target State.[23] In this context it must be conceded that according to some the damage inflicted must be severe.[24] If, however, there is no or merely minor material damage to the cyber infrastructure it is not really settled whether that activity can be considered a violation of territorial sovereignty.[25] The usual example given is espionage, including cyber espionage because international law lacks a prohibition of espionage. The fact that the data resident in the target system are modified by the act of intrusion is not considered sufficient to qualify it a prohibited violation of territorial sovereignty. It could, however, be argued that

18 Note that within the Exclusive Economic Zone and on the continental shelf coastal States do not enjoy territorial sovereignty but merely certain 'sovereign rights' with a view to the natural resources in those sea areas.

19 Vienna Convention on Diplomatic Relations, Article 27(1). Note that the computers and computer networks located in the diplomatic mission are protected by Article 22.

20 United Nations Convention on the Law of the Sea of 30 April 1982 (UN Doc. A/CONF. 62/122 of 7 October 1982), Articles 17 *et seq*., 37 *et seq*., 45, 52, and 53.

21 It is important to note that the prohibitions of the use of force and intervention only apply to States, i.e., to conduct attributable to a State. However, Article 51 of the UN Charter does not refer to the source of an armed attack. Today, there is general agreement that the right of self-defence also applies to armed attacks by non-State actors.

22 See, *inter alia*, R. Jennings/A. Watts (eds.), Oppenheim's International Law, Vol. I, para. 123, 9th ed., Jennings & Watts (eds.) (Harlow 1992).

23 *Ibid*., para. 119.

24 This is due to the fact that the use by a State of its territory very often causes negative effects on the territory of neighbouring States. Since the principle of territorial integrity is not considered to be absolute in character there are good reasons to maintain that damage below the threshold of severity must be tolerated and does not violate the territorial sovereignty (integrity) of the affected State.

25 Those who consider damage as relevant will not qualify such acts as violations of territorial sovereignty.

damage is irrelevant and that the mere fact that foreign State organs have intruded into the cyber infrastructure of another State is to be considered an exercise of jurisdiction on foreign territory that always constitutes a violation of the principle of territorial sovereignty.

According to the U.S. International Strategy for Cyberspace the following activities may qualify as violations of U.S. territorial sovereignty: attacks on networks, exploitation of networks, and other hostile acts in cyberspace that threaten peace and stability, civil liberties and privacy.[26] While the respective acts are not specified it seems that the U.S. government is advocating a rather wide scope of the principle of territorial sovereignty because it asserts the right to respond to such acts with all necessary means, including, if necessary, the use of (conventional) force.

As regards the cyber infrastructure thus protected by the principle of territorial sovereignty it is irrelevant whether it belongs to, or is operated by, governmental institutions, private entities or private individuals.

Moreover, such infrastructure is equally protected if it is located onboard aircraft, vessels or other platforms enjoying sovereign immunity. Article 95 LOSC[27] provides that "warships on the high seas have complete immunity from the jurisdiction of any State other than the flag State". According to Article 96 LOSC the same applies to "ships owned or operated by a State and used only for government non-commercial service". As regards state aircraft in international airspace there is general consensus that they enjoy sovereign immunity as well.[28] The Outer Space Treaty[29] and the Liability Convention[30] seem to justify the conclusion that space objects operated for non-commercial government purposes also enjoy sovereign immunity.[31] While there is no treaty rule explicitly according sovereign immunity to all objects used for non-commercial government purposes it is of importance that according to Article 5 of the UN Convention on State Immunity[32] a State enjoys immunity from the jurisdiction of the courts of another State with regard to its property.[33] This rule and the other rules just referred to give evidence of a general principle of public international law according to which objects owned by a State or used by that State for exclusively non-commercial government purposes are an integral part of the State's sovereignty and they are subject to the exclusive jurisdiction of that State if they are located outside the territory of another State. 'Sovereign immunity' means that any interference with an object enjoying such immunity constitutes a violation of the sovereignty of the State using the object for non-commercial government purposes.[34] It

26 International Strategy for Cyberspace, *supra* note 14, at 12 *et seq*.
27 *Supra* note 20.
28 See HPCR Manual on Air and Missile Warfare, Rule 1 (cc) and accompanying commentary, para. 6, available at http://ihlresearch.org/amw/Commentary%20on%20the%20HPCR%20Manual.pdf.
29 Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies of 10 October 1967, UN GA. Res. 2222 (XXII).
30 Convention on International Liability for Damage Caused by Space Objects of 1 September 1972, UN GA Res. 2777 (XXVI).
31 Note that space objects, such as satellites, used for governmental and commercial purposes, either by the State of registry or by that State in cooperation with a private corporation, do not enjoy sovereign immunity.
32 UN Convention on Jurisdictional Immunities of States and Their Property of 2 December 2004, UN GA Res. 59 (XXXVIII).
33 For an assessment see David P. Stewart, 'The UN Convention on Jurisdictional Immunities of States and Their Property', 99 AJIL 194-211, at 195 et seq. (2005).
34 For a first finding with regard to the sovereign immunity of warships see the Award of the Anglo-American Claims Commission, *The Jessie, The Thomas F. Bayard and The Pescawha*, Nielsen's Report, 479 *et seq*. (1926).

must be borne in mind, however, that in times of an international armed conflict the principle of sovereign immunity plays no role in the relations between the belligerent States. Then, objects enjoying sovereign immunity may be destroyed (if they qualify as lawful targets) or they are subject to seizure (booty of war)[35] by the respective enemy armed forces. Moreover, sovereign immunity is not limitless. For instance, the U.S. drone downed by Iran (allegedly by cyber means) had been in Iran's national airspace and it, thus, violated Iran's territorial sovereignty. Hence, Iran was entitled to use all necessary means, including cyber means, to terminate that violation.

Vehicles that do not serve exclusively non-commercial governmental purposes do not enjoy sovereign immunity. This, however, does not mean that they are not protected when located in areas or spaces that are not covered by the territorial sovereignty of any State. While they cannot be considered an integral component of a State's sovereignty, they are included into the protective scope of that sovereignty by the link of nationality. This means, that the respective State of nationality exercises exclusive jurisdiction over such objects when they are located on the high seas or in international airspace. Accordingly, any interference with such objects constitutes a violation of the sovereignty of the State of nationality (unless justified by a rule of public international law). This also applies to space objects. It is prohibited, under the Outer Space Treaty[36], to interfere with the activities of other States in the peaceful exploration and use of outer space. It is immaterial whether the space object is owned or operated by the government or by a private corporation. On the high seas and in international airspace the cyber infrastructure will regularly be located on board a vessel or aircraft and the determination of the State whose sovereignty and jurisdiction applies will depend on either the flag State principle[37] or on the national markings the aircraft carries.[38] That is different in outer space because satellites will in most cases have to be considered as qualifying as 'cyber infrastructure', i.e., without reference to a carrying platform. As in the case of aircraft, nationality of space objects is determined by registration.[39]

B. Exercise of Jurisdiction (Scope Ratione Materiae)

The second consequence of the applicability of the principle of territorial sovereignty to cyberspace is the wide-ranging right of the territorial State (including the flag State and the State of registry) to exercise its jurisdiction over cyber infrastructure and over cyber activities.

The concept of jurisdiction may be understood in a wide sense and referring to a State's "lawful power to act and hence to its power to decide whether and, if so, how to act, whether by legislative, executive or judicial means. In this sense, jurisdiction denominates primarily, but not exclusively, the lawful power to make and enforce rules".[40] As already noted above, the exercise of jurisdiction is not limited to a State's territory. For instance, a State exercises exclusive jurisdiction onboard vessels flying its flag and onboard aircraft registered in that State. Moreover, according to the principles of active and of passive nationality, a State is

35 See Yoram Dinstein, 'Booty of War', in: MPEPIL, available at http://www.mpepil.com.
36 *Supra* note 29.
37 Article 92 LOSC, *supra* note 20.
38 According to Article 17 of the Convention on International Civil Aviation (Chicago Convention) of 7 December 1944, "[a]ircraft have the nationality of the State in which they are registered".
39 See the Convention on Registration of Objects Launched into Outer Space of 15 September 1976, UN GA Res. 34/68.
40 Bernard H. Oxman, 'Jurisdiction of States', para. 1, in: MPEPIL, available at http://www.mpepil.com.

entitled to exercise its jurisdiction over the conduct of individuals that occurred outside its territory. Under the universality principle, the same holds true even if neither the perpetrator nor the victim are nationals of the State in question. Finally, the exercise of jurisdiction can be based upon the protective principle.[41]

For the purposes of this paper that deals with the principle of territorial sovereignty the forms of jurisdiction just referred to, although of importance in the cyber domain, need not be elaborated upon. Therefore, the focus will be on the scope of territorial jurisdiction.

It may be noted in this context that territorial jurisdiction does not necessarily presuppose territorial sovereignty. For instance, a State may exercise (exclusive) jurisdiction over territory leased or occupied.[42] It may also be noted that the jurisdiction conferred on coastal States in their Exclusive Economic Zone or on their continental shelf, although it may be conceived of as quasi-territorial in character, is only analogous to territorial jurisdiction *strictu sensu* because it is limited to certain activities.

For the purposes of this paper, it suffices to concentrate on a State's right to exercise its jurisdiction (i.e., to prescribe, enforce and adjudicate) over objects and persons physically (or legally) present in its territory. It seems to be undisputed that, unless limited by applicable rules of international law (probably including human rights law), cyber infrastructure located within the territory of a State and cyber activities occurring therein are susceptible to almost unlimited prescriptive and enforcement measures by the respective State. Territorial jurisdiction includes the right of a State to regulate, restrict or prohibit access to its cyber infrastructure either within its territory or from outside that territory. It must be re-emphasized that integration of physical components, i.e., of cyber infrastructure located within a State's territory, into the 'global domain' of cyberspace cannot be interpreted as a waiver of the exercise of territorial sovereignty and jurisdiction. In view of the mobility of users and of cloud or grid distributed systems it may very often be difficult to effectively exercise territorial jurisdiction. Still, those difficulties do not justify the conclusion that territorial jurisdiction, if applied to cyberspace, is but a 'toothless tiger'. To the contrary, States have regularly and quite successfully – while not always applauded – proven their willingness and determination to enforce their domestic law vis-à-vis all kinds of cyber activities.

A specific feature of territorial jurisdiction is the so-called 'effects doctrine' according to which a State is entitled to exercise its jurisdiction over a conduct that does not take place within its territory but that produces (harmful) effects in that territory.[43] A useful explanation of that doctrine has been provided by the European Attorney-General:

> "The two undisputed bases on which State jurisdiction is founded under international law are territoriality and nationality. The former confers jurisdiction on the State in which the person or the goods in question are situated or the event in question took place. The latter confers jurisdiction over nationals of the State concerned. Territoriality itself has given rise to two distinct principles of jurisdiction:

41 For a discussion of the different bases of jurisdiction see *ibid*., paras. 18 *et seq*.
42 *Ibid*., para. 15.
43 *Ibid*., paras. 22 *et seq*.

(i) subjective territoriality, which permits a State to deal with acts which are originated within its territory, even though completed abroad
(ii) objective territoriality, which conversely, permits a State to deal with acts which originated abroad but which were completed at least in part within its own territory.
[The effects doctrine] confers jurisdiction upon a State even if the conduct which produced [the effects] did not take place within the territory."[44]

Applied to the cyber domain, the effects doctrine may give rise to the exercise of jurisdiction over individuals who have conducted cyber operations against the cyber infrastructure in another State.[45]

In sum, it can be held that the principle of territorial sovereignty and the ensuing right of a State to exercise its territorial jurisdiction apply to cyberspace insofar as the cyber infrastructure within the territory (or on platforms over which the State exercises exclusive jurisdiction) is concerned. The same holds true for individuals present in that territory or for conduct that either takes place within that territory or that produces (harmful) effects thereon. The exercise of jurisdiction under any of the recognized bases under international law is limited only if there exist explicit rules to that effect. The characteristics of cyberspace do not pose an obstacle to the exercise of territorial sovereignty and jurisdiction.

5. OBLIGATIONS OF STATES IN CYBERSPACE AND THE ISSUE OF ATTRIBUTABILITY

A. Obligations of States in Cyberspace

This section does not deal with the entire spectrum of obligations States are to observe in cyberspace. Therefore, the prohibition of the use of force and the issue of 'armed attack' are not dealt with here. However, as noted above, the principle of territorial sovereignty does not only protect States by affording them exclusive rights but it also imposes obligations on them.[46] The protective scope of those obligations aims at the protection of the territorial sovereignty and integrity of other States.

Duty of Prevention

In view of its fundamental character the principle of (territorial) sovereignty entails an obligation imposed on all States to respect the (territorial) sovereignty of other States. As the ICJ held in the Nicaragua Case: "'Between independent States, respect for territorial sovereignty is an essential foundation of international relations' [...], and international law requires political integrity also to be respected."[47]

First of all, the obligation to respect the territorial sovereignty of other States applies to conduct that is attributable to a State. However, according to the Korfu Channel Judgment, respect for

44 ECJ, *Ahlström and others v. Commission (In re Wood Pulp Cartel)*, joint cases 89/85, 104/85, 114/85, 116-17/85 and 125-9/85, 96 ILR 148 et seq. (1994).

45 Hence, irrespective of the issue of attribution Estonia would be entitled to exercise its criminal and civil jurisdiction over those individuals who conducted the DDoS attacks against Estonian cyber infrastructure in 2007.

46 See the references *supra* note 5 and accompanying text.

47 ICJ, *Case concerning Military and Paramilitary Activities in and against Nicaragua (Merits)*, ICJ Rep. 1986, 14, at 106, para. 202, referring to its Judgment in the *Korfu Channel Case*, ICJ Rep. 1949, 35.

the territorial sovereignty of other States also implies the obligation of every State "not to allow knowingly its territory to be used for acts contrary to the rights of other States".[48] Accordingly, a State is required under international law to take appropriate acts in order to protect the interests of other States.[49] This obligation is not limited to "criminal acts"[50] but applies to all activities inflicting (severe) damage, or having the potential of inflicting such damage, on persons and objects protected by the (territorial) sovereignty of the target State.[51]

The duty of prevention, in the context of cyber attacks, has been correctly summarized as follows: "States have an affirmative duty to prevent cyberattacks from their territory against other states. This duty actually encompasses several smaller duties, to include […] prosecuting attackers, and, during the investigation and prosecution, cooperating with the victim-states of cyberattacks that originated from within their borders."[52]

It must be borne in mind that the term 'cyber attack' is often understood as comprising "remote intrusions into computer systems by individuals".[53] However, mere intrusions have to be excluded because they do not inflict direct (material) harm. Rather, intrusions must be considered acts of espionage that are not prohibited under public international law.[54] Since all States engage in espionage, including via the cyberspace, mere intrusions into foreign computers or networks are not covered by the prohibition.

The duty of prevention presupposes knowledge. This does not necessarily mean actual knowledge. The duty also applies to cases of presumptive knowledge. A State will have actual knowledge if its organs have detected a cyber attack originating from that State's territory or if that State has been informed by the victim State that a cyber attack has originated from its territory. Knowledge is to be presumed if the cyber attack can reasonably be considered to belong to a series of cyber attacks. It is important to note that the International Court of Justice has held that even if "a State on whose territory […] an act contrary to international law has occurred, may be called upon to give an explanation […] it cannot be concluded from the mere fact of the control exercised […] over its territory […] that that State necessarily knew, or ought to have known, of any unlawful act perpetrated therein".[55]

Hence, there are good reasons to conclude that the duty of prevention does not apply if the State from whose territory the respective acts have been committed has neither actual nor

48 ICJ, *The Korfu Channel Case (Merits)*, ICJ Rep. 1949, 1, at 22.

49 ICJ, *Case concerning United States Diplomatic and Consular Staff in Tehran*, ICJ Rep. 1980, 3, at 32 *et seq*., para. 68. See also Yoram Dinstein, *War, Aggression and Self-Defence*, at 206 (4th ed., Cambridge 2004).

50 Michael N. Schmitt, 'Preemptive Strategies in International Law', 24 Mich.J.Int'l.L., 513, at 540 *et seq*. (2003)

51 In the famous *Trail Smelter Case*, the Tribunal held *inter alia*: "This right [= sovereignty] excludes […] not only the usurpation and exercise of sovereign rights […] but also an actual encroachment which might prejudice the natural use of the territory and the free movement of its inhabitants. […] under the principles of international law […] no State has the right to use or permit the use of its territory in such a manner as to cause injury […] in or to the territory of another or the properties or persons therein, when the case is of serious consequence […]"; RIAA Vol. III, 1905, at 1963 *et seq*.

52 Matthew J. Sklerov, 'Solving the Dilemma of State Responses to Cyberattacks: A Justification for the Use of Active Defenses against States Who Neglect Their Duty to Prevent', 201 MLR 1-85, at 62 (Fall 2009).

53 *Ibid*., at 14.

54 See, *inter alia*, Schaap, *supra* note 6, at 139 *et seq*.

55 ICJ, *The Korfu Channel Case (Merits)*, ICJ Rep. 1949, 1, at 18.

presumptive knowledge. Such a conclusion is however, not necessarily generally accepted.

According to a position held in the literature the duty of prevention should be based on a State's "actions to prevent cyberattacks in general".[56] According to this position, "States that do not enact [stringent criminal laws and vigorous law enforcement] fail to live up with their duty to prevent cyberattacks. […] A state's passiveness and indifference toward cyberattacks make it a sanctuary state from where attackers can safely operate. When viewed in this light, a state can be held indirectly responsible for cyberattacks […]."[57] However, the mere theoretical possibility of a State that has not enacted criminal laws (and not being obliged to do so under an international treaty) becoming a sanctuary for cyber attackers is certainly not sufficient to justify the applicability of the duty of prevention.

There is, however, a situation that may be considered as sufficient for the assumption that the respective State had, or ought to have had, knowledge of the conduct. Such a situation may exist if a cyber attack has been launched from cyber infrastructure that is under exclusive government control and that is used for non-commercial government purposes only. Provided that the origin of, for instance, a cyber attack can be traced back to such government cyber infrastructure, there may at least be a rebuttable presumption that the respective State should have known of that use of its territory. It is important to note that a rebuttable presumption of knowledge does not mean that the respective conduct is, thus, attributable to the State. That would mean that the aggrieved State would be entitled to resort to counter-measures, including, where applicable, to the use of force. However, the rebuttable presumption as such is not sufficient to either attribute the conduct to the State or to serve as a legal basis for counter-measures although that might be the case with a view to events occurring in the physical world. In cyberspace such an approach could lead to an escalation and it would certainly impose on States too far-reaching obligations because the government cyber infrastructure may have been usurped by another State or by non-State actors, such as terrorists or other criminals.

Without prejudice to the problem of attributing a cyber attack to a State it may not be left out of consideration that the duty of prevention applies only if and insofar as the cyber attack has been launched from the territory of a given State. Despite of the complexity of cyberspace some might be inclined to recognize the duty of prevention to apply also to cyber attacks/cyber operations that are routed through the cyber infrastructure of another State. It is, however, unsettled whether the transit of data through another State brings into operation the obligation of prevention even if the transit State knows, or should have known, of the use of the cyber infrastructure located on its territory. On the one hand, the respective data may only be parts of a data packet. While the packet as such may be considered a 'cyber weapon', its constituent parts may be transmitted over different nodes. On the other hand, in most cases it would be meaningless to oblige the transit State to take preventive action because the data may be rerouted and may therefore nevertheless arrive at their destination.

Further Obligations

Finally, it may be added that State practice seems to justify the conclusion that there is a growing readiness of States to accept obligations that are of a more general character than the obligation to refrain from harmful conduct or to prevent such conduct.

56 Sklerov, *supra* note 52, at 71.
57 *Ibid*.

For instance, the U.S President has taken the position that identifying the rules and principles of international law applicable to cyberspace must be guided by the "broad expectations of peaceful and just interstate conduct to cyberspace."[58] The U.S. President emphasizes that States "need to recognize the international implications of their technical decisions, and act with respect for one another's networks and the broader Internet"[59] and he demands that the emerging norms are guided by five criteria, including global interoperability, network stability and cyber security due diligence.[60] Indeed, global interoperability is one of the main characteristics of the Internet and it can only be preserved if "States […] act within their authorities to help ensure the end-to-end interoperability of an Internet accessible to all". Network stability presupposes that States do not "arbitrarily interfere with internationally interconnected infrastructure". Since cyber security due diligence is understood to imply that "States should recognize and act on their responsibility to protect information infrastructures and secure national systems from damage or misuse", it may be considered as already being part of customary international law, i.e., reflective of the obligation of prevention discussed above. According to the position taken here the criteria enumerated in the International Strategy for Cyberspace may not yet have the status of customary international law but they may well be accepted by a considerable number of States – at least by those that are 'like-minded'. The criteria may in any event be considered as being of a potentially norm-creating character, thus contributing to the progressive development of customary international law.

B. Attributability

An effective protection of territorial sovereignty in the cyber domain presupposes that a given conduct can be attributed to another State. Of course, the rather strict criteria of attributability in Articles 4 to 11 of the ILC's Draft Articles on State Responsibility[61] are designed for the purpose of State responsibility and they do not necessarily preclude the application of more liberal criteria with a view to determining the origin of a cyber attack. It is, however, unclear whether States are prepared to agree on such criteria.

It is generally agreed that, in view of the architecture and characteristics of cyberspace, it is "virtually impossible to attribute a cyberattack during an attack. Although states can trace the cyberattack back to a computer server in another state, conclusively ascertaining the identity of the attacker requires an intensive, time-consuming investigation with assistance from the state of origin."[62] The cyber attacks on Estonia (2007) and on Georgia (2008) prove the correctness of this finding. The U.S. Department of Defense (DoD) has also stressed that the "often low cost of developing malicious code and the high number and variety of actors in cyberspace make the discovery and tracking of malicious cyber tools difficult. Most of the technology used in this context is inherently dual-use, and even software might be minimally repurposed for malicious action." In conclusion, the DoD admits that the "interconnected nature of cyberspace poses significant challenges for applying some of the legal frameworks developed for specific physical domains."[63]

58 International Strategy for Cyberspace, *supra* note 14, at 9.
59 *Ibid*., at 10.
60 *Ibid*. The remaining two are reliable access and multi-stakeholder governance.
61 Draft Articles on Responsibility of States for Internationally Wrongful Acts, UN GA Res. 56/82 of 12 December 2001.
62 Sklerov, *supra* note 52, at 7.
63 Cyberspace Policy Report, *supra* note 14, at 8.

Despite of the difficulty of verifying the location from which an attack was launched or of identifying the attacker, the DoD has announced it would "actively seek to limit the ability of such potential actors to exploit or attack the United States anonymously"[64] It is, of course, almost a commonplace that inter-agency and international cooperation as well as information sharing are a necessary prerequisite to achieve that goal. In view of the special characteristics of cyberspace it may well be stated that international law provides an obligation to cooperate if States are prepared to take measures in cyberspace. It will be interesting to see whether the DoD's efforts to "assess the identity of the attacker via behavior-based algorithms" and to "significantly improve its cyber forensics capabilities"[65] will be successful and, what is equally important, accepted by other States as conclusive or sufficient evidence of the source of a cyber attack.

6. CONCLUSIONS

Territorial sovereignty has proven to be a powerful and effective norm of international law that can be applied to cyberspace without far-reaching modifications if cyberspace is understood as comprising components – or: cyber infrastructure – that is located in a State's territory or otherwise protected by the principle of territorial sovereignty. It may not be forgotten that this finding does not imply that all aspects of the protection of territorial sovereignty have thus been clarified. For instance, there still is no consensus among States as to which cyber operations qualify as a prohibited use of force according to Article 2 (4) of the UN Charter or as an armed attack under Article 51. The rather abstract references to 'critical infrastructure' are not very helpful if there is no consensus as to which objects and institutions are to be considered 'critical' in nature.

Equally effective is the concept of territorial jurisdiction. Accordingly, States are entitled to regulate cyber activities occurring within their territories and to enforce their domestic law. Although States enjoy an almost unlimited right to exercise their territorial jurisdiction with a view to cyber activities and cyber infrastructure within their territories there is an undisputable need for an internationally agreed understanding that the Internet's functionality and thus the benefits it entails would be seriously challenged if States do not exercise their territorial jurisdiction "with respect for one another's networks and the broader Internet".[66] Therefore, the five criteria identified by the U.S. President in the International Strategy for Cyberspace should be taken up by other governments. They are of a potentially norm-creating character and they would assist in a clarification of the existing rules and principles of international law that apply to the cyber domain.

Finally, governments should cooperate with a view to improving their capabilities in the area of cyber forensics. Such cooperative efforts are necessary not only in order to identify attackers but also for a more effective deterrence of malevolent States and non-State actors.

64 *Ibid*., at 4 *et seq*.
65 *Ibid*.
66 International Strategy for Cyberspace, *supra* note 14, at 10.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

When "Not My Problem" Isn't Enough: Political Neutrality and National Responsibility in Cyber Conflict

Jason Healey
Cyber Statecraft Initiative
Atlantic Council
Washington, D.C., U.S.A.
jhealey@acus.org

Abstract: Cyber conflict may not be new, but it is far from old. And as with any other major, disruptive global trend, there are vexing questions on which traditional international norms still apply, whether they apply but with modifications, or whether entirely new norms must be invented. One of the most important norms has been for states to be able to remain neutral in response to international conflict, with rights and responsibilities guaranteed by the Hague Convention. Because of the nature of cyber conflict, such legal norm may be less useful than a modified norm of *political* neutrality. The Internet protocols themselves route cyber attacks through any number of neutral countries, cyber conflicts are usually not so destructive to obviously trigger international law, and the identity or nationality of the belligerents may not be obvious.

Nations might (and probably *should*) accordingly come under political pressure to take reasonable steps to stop cyber attacks, regardless of whether or not it is a formal treaty obligation. This paper explores this issue and ways a nation may be less than neutral, tying this to a ten-point spectrum of state responsibility to help determine just how responsible a nation might be in a cyber conflict. To illustrate potential new norms in action, the paper then describes a notional cyber conflict which shows how the nations' rights and responsibilities are influenced by the four factors of severity, obviousness, "stoppability," and duration. The paper concludes with a short section on the commercial neutrality during cyber-conflict, given the critical role that the private sector has played in the creation and operation of cyberspace.

Keywords: *neutrality, cyber conflict, national responsibility, Hague Convention, Law of Armed Conflict, political neutrality, commercial neutrality*

1. INTRODUCTION

Since cyberspace makes us all neighbors, more nations are likely to be affected by conflicts in cyberspace than in the air, land or sea. These nations will have to take more active steps to stop attack traffic if they wish to remain neutral.

Nations are increasingly looking to limit future conflicts, to bring these under more control, just as more traditional wars were restrained through treaties, conventions and norms. But it is still unknown how well the old agreements will hold up and what must be reinvented because of the nature of cyberspace and cyber conflict.

One of the most important norms has been for states to be able to remain neutral in response to international conflict, with rights and responsibilities guaranteed by the Hague Convention. Because of the nature of cyber conflict, such legal norm may be less useful than a modified norm of *political* neutrality. The Internet protocols themselves route cyber attacks through any number of neutral countries, cyber conflicts are usually not so destructive to obviously trigger international law, and the identity or nationality of belligerents may not be obvious.

Nations might (and probably *should*) accordingly come under political pressure to take reasonable steps to stop cyber attacks, regardless of whether or not it is a formal treaty obligation. This paper examines one aspect of this, political neutrality in cyber conflict. New norms will develop as "not my problem" will no longer be acceptable.

This paper will start the examination of political neutrality with a literature review of neutrality and cyber conflict, especially the legal aspects which features in most of the literature. However, after this introductory section, the paper shifts from legal to political neutrality, which allows more flexibility to adapt to the nature of cyber conflict. After this, the paper moves on to specific ways a nation could be less than neutral, tied to a ten-point spectrum to help understand responsibility and neutrality. A notional example of a cyber conflict illustrates how political neutrality might work in practice and highlights four factors likely to influence political neutrality – severity, obviousness, "stoppability," and duration – and areas for further research.

2. CYBER CONFLICT AND NEUTRALITY: HOW DID WE GET HERE?

The obvious starting point in this discussion is "what is meant by neutrality?" Though the concept is an old one, the current legal international concept was codified in the Hague Convention of 1907, which discusses rights and duties, and begins as clearly as possible, "The territory of neutral Powers is inviolable." A definition that seems to be widely used is one from the dictionary published by the U.S. Department of Defense (DoD). Neutrality here is defined as in international law, the attitude of impartiality during periods of war adopted by third states toward a belligerent and subsequently recognized by the belligerent, which creates rights and duties between the impartial states and the belligerent.[1]

1 JP 1-02, "DoD Dictionary of Military and Associated Terms", January 2012, p. 234.

This definition lacks mention of neutrality in cyber conflict, but this is no surprise as it does not discuss the obvious ways neutrality differs in the other domains of land, air, sea or space either. The U.S. government has been very clear that it will treat cyberspace as it does these other domains, not least for the applicability of international law.

The White House International Strategy for Cyberspace declared that "Consistent with the United Nations Charter, states have an inherent right to self-defense that may be triggered by certain aggressive acts in cyberspace."[2] Similarly, the commander of Cyber Command in testimony to Congress declared, "all military operations must be in compliance with the laws of armed conflict—this includes cyber operations as well. The law of war principles of military necessity, proportionality and distinction will apply [...]"[3] If needed, one provision of the 1934 Communications Act allows the President to close down communications stations and remove equipment if needed, to "in order to preserve the neutrality of the United States."[4]

Most of this recent attention, however, has been focused only on two areas: how the United States would respond to an attack on itself (or its allies) and how the laws of armed conflict (LOAC, also known as International Humanitarian Law or IHL) apply to offensive military operations. There has been little or no mention of how neutrality applies to cyber other than an implication it would be handled similar to any other domain. This is not straightforward, of course.

The only official U.S. document that goes into any depth on neutrality in cyber conflict is a 1999 document from the DoD General Counsel, *An Assessment of International Legal Issues in Information Operations*.[5] This early paper covered an impressive range of issues relating to cyber operations (though they were not then called by that term) including neutrality and "self-defense in neutral territory." This paper made several important contributions, including making it clear that

- "If a neutral nation permits its information systems to be used by the military forces of one of the belligerents, the other belligerent generally has a right to demand that it stop doing so."
- "A neutral Power is not called upon to forbid or restrict [communications], so long as such facilities are provided impartially to both belligerents."
- The use of a "nation's communications networks as a conduit for an electronic attack would not be a violation of its sovereignty in the same way that would be a flight through its airspace by a military aircraft."
- Nations need not have much concern "for the reaction of nations through whose territory or communications systems a destructive message may be routed."
- "Transited state would have somewhat more right to complain if the attacking state obtained unauthorized entry into its computer systems as part of the communications path to the target computer."

2 White House, International Strategy for Cyberspace, 2011, p. 14.
3 General Keith Alexander, Advance Questions for Lieutenant General Keith Alexander, USA Nominee for Commander, United States Cyber Command, 2010, p. 15.
4 Communications Act of 1934, Section 606c.
5 Department of Defense General Counsel, *An Assessment of International Legal Issues in Information Operations, May 1999*.

The general US approach, to treat cyberspace as similar to other domains, is supported by material from the International Committee of the Red Cross, whose work is based on the Geneva and Hague Conventions that are a foundation for LOAC. An official paper from 2004 by Knut Dörmann, Deputy Head of the ICRC Legal Division, argues that under the Geneva Convention (and its Additional Protocols, signed but not ratified by the United States), "the fact that a particular military activity constituting a method of warfare is not specifically regulated, does not mean that it can be used without restrictions."[6] This paper discusses many ways that LOAC would apply to cyber operations, but includes little on neutrality. Andrew Carswell, an armed forces delegate to the ICRC, has gone farther to describe their view on neutrality in a 2011 presentation. Starting with an explanation of the Hague Convention laws (and a sense they have a "slightly musty quality") he examines several scenarios on how neutrality might apply to cyber conflict.[7]

Neutrality in cyber conflict is vexed by any number of challenging questions, such as these, from a paper by Sean Kanuck, now a senior U.S. intelligence official:

1. "What if a neutral party did not know when its sovereignty was breached to conduct an attack or was technically incapable of restricting belligerents' use of its […] networks without irreparably harming its own governmental functions or economy?
2. "What if the tools required to conduct or defend against a cyber attack needed to be pre-positioned in global networks to be most efficacious?
3. "What if a sovereign did not exercise due diligence in preventing its own subjects from criminally compromising foreign computer systems and later using them to attack a third sovereign nation?"[8]

To help the discussion move past theoretical questions, two military officers from a U.S. military cyber defense unit took the discussion in a very practical direction. Stephen Korns and Joshua Kastenburg examined one of the most important international cyber conflicts, the Russian invasion of Georgia in 2008, when a U.S. internet service provider hosted the website of the Georgian president, with important implications for America's role as a neutral or belligerent. Korns and Kastenburg, as one of the few full-length treatments on the subject provide an excellent definition of legal cyber neutrality:

> "Cyber neutrality, therefore, is the right of any nation to maintain relations with all parties engaged in a cyber conflict. Under a traditional international law rubric, to remain neutral in a cyber conflict a nation cannot originate a cyber attack, and it also has to take action to prevent a cyber attack from transiting its Internet nodes."[9]

6 Knut Dörmann, "Applicability of the Additional Protocols to Computer Network Attacks," 2004, p. 2, *available at* http://www.icrc.org/eng/resources/documents/misc/68lg92.htm.

7 Andrew Carswell, "Neutrality in Cyberwar," Presentation To The Internet In Bello: Seminar On Cyber War, Ethics & Policy, UC Berkeley School of Law, 2011, *available at* http://www.law.berkeley.edu/files/Neutrality_in_Cyber_War_for_web.pdf.

8 Sean Kanuck, "Sovereign Discourse on Cyber Conflict Under International Law," Texas Law Review, Vol.88, Issue 7, 2010, p. 1593.

9 Stephen W. Korns And Joshua E. Kastenberg, "Georgia's Cyber Left Hook," Parameters, Winter 2008-2009, p. 62.

Korns and Kastenburg also highlight an important additional aspect that is rarely mentioned in other works, the role of the private sector, which dominates in cyberspace in a way they do not in other domains, with important implications for neutrality. According to their paper,

> "Private industry owns and operates the majority of the Internet system. During a cyber conflict, the unregulated actions of third-party actors have the potential of unintentionally impacting US cyber policy, including cyber neutrality. There is little, if any, modern legal precedent."[10]

Kastenburg later wrote a follow-up article in a U.S. Air Force Law Review that also examined this incident but with a more legal perspective.[11]

This focus on real-world events marks an important trend in the literature, an increasing focus not on the *legal* implications of neutrality, but the *political* importance. After all, nations can still insist other nations take actions to mitigate the effects of a cyber conflict, even if international lawyers are still parsing over "musty" treaties and arguing over the meanings.

This expectation that nations have some positive obligation to assist during cyber conflicts to which they are not a belligerent is tied to the ideas of national responsibility or sovereignty and has been explored in the writings of Sean Kanuck (already referenced above) along with David Graham ("Cyber Threats and the Law of War" in the *Journal of National Security Law and Policy*, 2010) and Patrick Franzese ("Sovereignty in Cyberspace" in *Air Force Law Review*, 2009). These authors all have general consensus around certain points, such as (in Franzese's words), "Many of the designers of cyberspace viewed it as an intellectual nirvana free from the constraints of the 'real' world. In reality, however, cyberspace is part of the 'real' world and thus subject to its constraints and order—in other words, subject to state sovereignty."

More recently, a paper by this author explores the idea further and describes a ten point spectrum of national responsibility.[12] The present paper will apply and extend this spectrum to bring clarity and rigor to the idea of political neutrality in cyber conflict.

3. WHAT DO WE MEAN BY POLITICAL, VICE LEGAL, NEUTRALITY IN CYBER CONFLICT?

Even in the traditional domains of air, land, sea it may not be clear how to apply the Hague guarantee that "The territory of neutral Powers is inviolable." But in those domains neutrality is far clearer than in cyberspace.

The Internet protocols themselves route cyber attacks through any number of neutral countries in ways that may not be known – or even predictable – by a belligerent. Moreover, the cyber conflicts seen so far are typically criminal intrusions, criminal denial of service attacks, nuisance

10 *Ibid*., p. 1.

11 Joshua E. Kastenberg, "On-Intervention And Neutrality In Cyberspace: An Emerging Principle In The National Practice Of International Law," Air Force Law Review, Volume 64, 2009.

12 Jason Healey, "Beyond Attribution: Seeking National Responsibility in Cyberspace," Atlantic Council, 2012. Earlier published as "The Spectrum of National Responsibility for Cyberattacks" in the Brown Journal of World Affairs, 18.1 Fall/Winter 2011.

attacks by bored or aggressive hackers, or espionage. None of these obviously rise to the level of "armed conflict" or other thresholds required for most international laws on conflict to apply. Even in conflicts with clear national security implications (such as Estonia in 2007 and Georgia 2008) the disruption caused was short-term, reversible, and did not appear to have caused any casualties. Lastly, the identity or nationality of the belligerents may not be obvious. Indeed, the target of an attack may not even know they are under attack.

All of this makes a strict legal approach, bound to existing treaties, problematic. Even more problematic would be attempting to modify existing treaties. So far the world has only seen a subset of the likely kinds of cyber conflict. Modifying treaties to accommodate only those we have seen so far would be myopic and modifying them to include conflicts we have not yet seen, and can only imagine, would be folly.

Political neutrality fills this gap especially as it can operate under the strict legal thresholds and be always applicable. For example, Russia is under no legal obligations to be impartial between the belligerents in the Syrian uprising, since it is not an international armed conflict. Despite this lack of legal standing, other nations can apply the political (that is, diplomatic) pressure of moral condemnation to convince Russia to cease shipping weapons to the Assad regime.

In contrast to the more strictly defined legal norms of the Geneva and Hague Conventions, political neutrality allows a wider range of expectations and responses. Since it is judged, not by international tribunals, but heads of state and public opinion it establishes in essence a separate set of norms for international behavior.

The attacks against Estonia in 2007 provide a practical example of political neutrality in cyber conflict. Cyber attacks inundated Estonia during a political crisis between Estonia and Russia. The attackers followed "instructions provided on Russian-language Internet forums and websites," and were supported by comments from senior Russian politicians.[13] The attacks themselves appeared to originate from – or were routed through – 178 different countries. All of these countries which were asked, bar one, agreed to help cease the attacks and assist the Estonian investigation. The exception was Russia, which waited six weeks (indeed, after the conflict was over) to refusal, an act that "was not the inevitable legal solution, considering both earlier [Estonian] cooperation practice with Russia and the practice with other countries with whom identically phrased bilateral agreements."[14] Ever since, Russia has been presumed to have been, if not a legally defined belligerent, then at least complicit and iniquitous.

This then, is the heart of the political neutrality in a cyber conflict. Some nations certainly helped Estonia not to be impartial, but rather the opposite, to give them active assistance in the face of perceived bullying. Other nations, however, probably did indeed seek impartiality, choosing not to be a source of attack traffic tormenting a fellow nation during a crisis to which they were not a party.

Nations might (and probably *should*) come under political pressure to take reasonable steps to stop cyber attacks, regardless of whether or not it is a formal treaty obligation.

13 Eneken Tikk, et al, "International Cyber Incidents: Legal Considerations," NATO CCDCOE, 2010, p. 33.
14 *Ibid*., p. 27.

4. HOW CAN A NATION BE LESS THAN NEUTRAL IN A CYBER CONFLICT?

Most legal literature on neutrality and cyber conflict focuses on a single issue: "Does routing of attacks by a belligerent state through the internet nodes of a neutral country violate its neutrality?" as it was put by the ICRC.[15] This is perhaps the wrong perspective, given the kinds of cyber conflict to date, as embodied in the 2007 attacks on Estonia.

A better phrasing may be "During a conflict, what obligations does a State have to stop attacks coming from its territory or citizens?" This similar, but broader, question encompasses the possibilities that a State will still have responsibilities not only when a belligerent routes traffic through its "internet nodes."

During the Estonia crisis, most attacks were not "routed" as such through those 178 nations in the way we normally think of a weapon system being routed. These attacks were not predominantly cyber missiles, launched from one the government of one belligerent and passing through the territory of other nations on its way to the target. Rather, most of the 178 nations would have either (1) hosted infected computers (called bots or zombies) that were under the control of non-state actors in one belligerent country, or (2) been the location from which non-state patriot hackers launched such attacks in support of their original motherland.

Indeed, though being the source of attack traffic is the most visible way that nations can lose their political neutrality in a cyber conflict, it is not the only way. Here is a more inclusive, but still partial, list:

1. Hosting bots in its physical territory.
2. Hosting command and control nodes of a network of bots (i.e., a botnet).
3. Attacks pass through physical territory on their way to the target.
4. Residents in its physical territory are participating in the attack.
5. Hosting legitimate military or dual-use targets of interest to one of the belligerents.
6. Hosting chat rooms that are coordinating the attack.
7. Senior leaders are encouraging attacks.
8. Refusing to respond to requests for help.

For a State to consider itself strongly neutral, it should be working to mitigate all of these symptoms of partiality – many of which fall under other obligations, such as the Council of Europe´s Convention on Cybercrime of 2001 (Budapest Convention).

Note that, importantly, this flips the legal norm on its head. Because attacks are internationally routed in ways that may not be knowable to an attacker, the traditional norm based on a responsible on the attacking belligerent becomes highly problematic, at times nonsensical. Some of this responsibility must be picked up by nations along that attack path to take reasonable steps to mitigate the attack if they can.

15 Andrew Carswell, "Neutrality in Cyberwar," Presentation To The Internet In Bello: Seminar On Cyber War, Ethics & Policy, UC Berkeley School of Law, 2011, *available at* http://www.law.berkeley.edu/files/Neutrality_in_Cyber_War_for_web.pdf.

5. IN CONTEXT: AN EXAMPLE OF CYBER CONFLICT

To help pull apart these threads of political neutrality, the following example gives a realistic conflict scenario.

Phase 1: Zendia directs its hacker groups to deface and disrupt webpages of the Ruritanian leadership and the networks of banks, utilities and online stores. The botnets used in the attack come predominantly from five countries: Zendia, Trissalia, Floria, Pollabia, and Glospland. The attacks cause no casualties or significant disruption, though they are inconvenient. In response, Ruritania asks for assistance. Zendia and its client Trissalia unsurprisingly refuse to take any action; Floria attempts to stop the attacks but cannot, lacking technical and law enforcement capacity. Pollabia and Glospland are able to stop the attacks.

After the attacks continue for some weeks, pro-Ruritanian hackers both in that country and the diaspora, organize a sizable counteroffensive against Zendia using botnets in all the above countries. Ruritania asks these attacks to stop as they are "not helpful" to de-escalate the situation. Zendia requests help and again Floria tries to help but cannot. Trissalia, which had claimed it was unable to track down the hackers or computers involved in the operation against Ruritania, suddenly finds the ability to help Zendia. The attacks are rapidly stopped and Trissalia extradites those responsible to a gloomy fate in Zendia. Pollabia stops these attacks as effectively as it did for those against Ruritania. Glospland responds to the requests from Zendia, but still sends technical teams to Ruritania to bolster their defenses and provides emergency loans to buy advanced security kit.

In addition to formally making *demarches* to the unhelpful countries, Ruritania protests formally in regional security forums and at the United Nations Security Council and General Assembly.

Phase 2: Since Ruritania's defenses have become significantly better at blocking attack traffic, Zendia sends teams to both Trissalia and Floria to build additional attack infrastructure and enlist other hackers. Now, these countries are not just the source of botnet traffic, they have Zendian hackers conducting attacks from their own soil. In addition, Zendia has initiated a new line of attack. Rather than massive (and noticeable) denial of service attacks using botnets, they begin "low and slow" intrusions, routed through all the countries involved. These are hard to detect, even by watchful defenders using advanced gear.

Ruritania feels that Trissalia and Floria, with attack teams on their own soil, these countries have far stronger responsibilities now that their role in the crisis is more direct. Unfortunately, the Florian government is still unable to stop the attacks and Trissalia unwilling. It asks for help to stop the "low and slow" attacks, but as these are so difficult, it does not complain when little help is forthcoming.

Phase 3: The attacks ratchet up: nearly 200 people have been left dead and injured after the disruption of traffic lights, medical records, and local electrical power. Floria, which had been unable to stop the attacks earlier, realize the change in the nature of the conflict and are able to implement a heavy handed, but effective stop to the attacks from their territory. The heads

of state of Floria, Pollabia, and Glospland come together to demand first that Zendia cease to use their territory in the onslaught against Ruritania and threaten a response. Some of their more academic-minded international lawyers resist, saying there is far from a clear cut case that the Zendian leadership is truly responsible and, even if they were, the law is far from clear unless the UN Security Council acts. Glospland goes further, saying the attacks must stop, from wherever their source, or else there will be a military response. In the meantime, they implement sanctions, use their diplomats and political leaders to vilify Zendia and use other levers of power.

6. UNDERSTANDING POLITICAL NEUTRALITY IN CYBER CONFLICTS

As noted in an earlier previous section and illustrated by this example, there are many ways a nation can be less than neutral in a cyber conflict. Accordingly, this means there are many shades of responsibility each nation can bear but, as yet, there has not been any easy way to categorize these. To understand this example, the Spectrum of State Responsibility[16] (see Table 1) is helpful – but not conclusive – to determine how each neutral a nation really is. This spectrum assigns ten categories, each marked by a different degree of responsibility, based on whether a nation ignores, abets, or conducts an attack. The spectrum starts from a very passive responsibility—a nation having insecure systems that lead to an attack—up to very active responsibility—a national government actually planning and executing an attack. Countries that fall into the first two categories ("State Prohibited" and "State Prohibited But Inadequate") have only very passive responsibility – and are the most politically neutral – since they will, at the least, attempt to cease any participation in the attacks. In the next four categories ("State Ignored," "State Encouraged," "State Shaped," and "State Coordinated") the nation is in no sense neutral, as it is actively ignoring or abetting the attacks. In the final four categories ("State Ordered," "State Rogue Conducted," State Executed," and "State Integrated"), the state has a much more direct hand as a belligerent, either ordering attacks or conducting them itself.

TABLE 1: THE SPECTRUM OF STATE RESPONSIBILITY

1. **State-prohibited.** The national government will help stop the third-party attack.
2. **State-prohibited-but-inadequate.** The national government is cooperative but unable to stop the third-party attack.
3. **State-ignored.** The national government knows about the third-party attacks but is unwilling to take any official action.
4. **State-encouraged.** Third parties control and conduct the attack, but the national government encourages them as a matter of policy.
5. **State-shaped.** Third parties control and conduct the attack, but the state provides some support.
6. **State-coordinated.** The national government coordinates third-party attackers such as by "suggesting" operational details.
7. **State-ordered.** The national government directs third-party proxies to conduct the attack on its behalf.
8. **State-rogue-conducted.** Out-of-control elements of cyber forces of the national government conduct the attack.
9. **State-executed.** The national government conducts the attack using cyber forces under their direct control.
10. **State-integrated.** The national government attacks using integrated third-party proxies and government cyber forces.

The spectrum can be used both to describe individual attacks or a campaign of related attacks, and is meant to be both for the operational cyber defenders (*"General, this attack against us is probably* state-ordered. *If we ask that nation for cooperation, they*

16 For more details, see previous cite for Healey, "Beyond Attribution: Seeking National Responsibility in Cyberspace", *supra* note 12.

will not help us, and we will tip our hand.") and the policy community (*"The policy of our nation is to hold nations accountable for any* state-ordered *attacks as if those attacks were coming from the uniformed military services. You can't hide behind proxies."*).

The Spectrum of State Responsibility provides a much clearer vocabulary for political neutrality. Nations at the high end of the spectrum have more characteristics of a belligerent while those at the bottom end are the most neutral. Nations that take direct actions for one belligerent but not all of them, may be seen as helpful but not neutral.

How politically neutral are each of the five countries in the earlier example? **Zendia** proved itself as not at all neutral. Indeed, it should be considered a belligerent, as it actually "ordered" the attacks (rather than merely ignoring, encouraging, shaping or coordinating them), putting it at level 7 in the spectrum. **Ruritania** was also a belligerent, in that there were broad societal attacks, but it did try to rein in counterattacks. **Trissalia** did not order any attacks but clearly provided all support to one side, the Zendians, and ignored requests from the other party. This means it is at least at level 3 of ignoring the attacks. **Floria** and **Pollabia** responded neutrally to both parties, though the former's response was feckless, putting these countries at levels 2 and 1 respectively. **Glospland** acted neutrally in stopping the attacks, putting it at level 1, but did later support Zendia as the party facing the online aggression.

At no point was "attribution" particularly important: indeed the applicable norms would prohibit supporting a conflict even if none of the belligerents are known. The attacks do not need to be traced to determine the computers and command and control network involved, then the people and organizations that were ultimately in control. The obvious attack traffic could have just been stopped, regardless of the geopolitical situation.

In the scenario, the technical community would try bottom-up technical attribution, but top-down attribution, would clearly point to Zendia as being to blame. The Zendian government would certainly try to hide behind the fiction that their involvement could not be "proved" but especially once there were casualties, this cover would have become increasingly threadbare.

As the scenario proceeded, though the spectrum remained helpful, there were obviously other factors in play. The most important of these are the overlapping criteria of severity, obviousness, "stoppability," and duration.[17]

- **Severity**: Some conflicts are more dangerous than others; the more intense and deadly the stronger the requirement for positive actions to remain neutral.
- **Obviousness**: Some attack patterns are far more evident which implies a stronger responsibility for a nation to not allow them if they want to remain neutral.
- **Stoppability**: Some attack patterns are far easier to restrict which implies a stronger responsibility for a nation to not allow them.
- **Duration**: The longer the cyber conflict, the stronger the need for a country to take actions to remain neutral. A single attack packet that passes through the nation's system deserves less response than a campaign lasting months.

17 Note these are related to, but not identical to the "scope, duration and intensity" test for whether an attack reaches the threshold of "armed attack" in the UN Charter (see Thomas Wingfield and others).

These important points often seem undervalued or even ignored in the current discussion which often focus on today's headlines on cyber crime and espionage – which are important but not severe. Accordingly, the norms of politically neutrality seem hard to find and weak. Yet they are not only realistic but help to give far more clarity on the appropriate norms. Once there is a more severe crisis with casualties and real damage, political neutrality will become more important. In the same way, discussion on political neutrality must distinguish between attacks which are the most easily detected and stopped, as there is a higher obligation to stop these.

In the example above, Floria did not have the capacity to be as politically neutral as it would have liked. But it turned out this incapacity was conditional, and lasted only as long as the attacks were a crisis but not a catastrophe. Once there were hundreds of casualties, however, it felt a moral obligation (and probably a responsibility both to international and domestic audiences) to make strenuous efforts.

In the earliest phase, Ruritania was disappointed with the nations that failed to stop the attacks, especially those nations that did not even try. One reason was that denial of service attacks and botnets are fairly easy to both spot and stop. Internet Service Providers (and by extension, States) can typically spot this kind of traffic transiting their systems and there are methods to counter them. Ruritania was right to be upset by nations that could not reign in these attacks. By the later phases, some of the attacks had become "low and slow" and Ruritania no longer had such a high expectation.

As for duration, this notional example is far closer to the history of actual cyber conflicts, which are not won or lost "at the speed of light" as is often imagined. Though individual engagements can indeed be that quick, the conflicts themselves are usually months-long campaigns with repeated clashes.

7. COMMERCIAL NEUTRALITY

The dominant difference between conflict in cyberspace is not the speed of operations, nor the fuzziness of borders, or global reach. While important, these are dwarfed by the fact cyberspace is owned and operated overwhelmingly by the private sector. Any relevant national-security relevant conflict will be fought in the networks and systems of individual companies which built them for their own purposes and which may decide they want nothing to do with the conflicts of their host nations.

For example, imagine if there were a repeat of the 2007 attacks against Estonia. Microsoft, McAfee, Symantec, Kaspersky and other companies may want to be seen as neutral, providing impartial service to both belligerents. They may not be able to, however, either because of a governments order or because one side sees them as being a tool of, or disproportionately helping, the other.

Indeed, commercial pressures already enforce something very much like commercial neutrality. Bill Woodcock of the Packet Clearing House describes the long track record of successful

cooperation between the world's largest network providers to stop the most disruptive attacks.[18] He describes a common scenario where one provider, say in the United States, may see a massive attack coming from their connection from an Internet exchange point in, for example, London. These major providers have a special authenticated hotline system for the U.S. downstream provider to contact the upstream provider in London to ask them to stop the attack streams, since they are just being dropped by the US provider. This is usually in everyone's interest, since the upstream provider is paying to send this traffic which will never be delivered, taking up their bandwidth in the meantime. Why pay to send bits that will never be delivered? Indeed, it is then in the downstream provider's interest to ask for a cessation of attack traffic from whatever provider is sending into them, who can continue this chain to the originating network owner.

This process is not being done for any reasons related to 'neutrality,' certainly not because of any articles of the Hague Convention. They do it because it is cheaper, more efficient, and just good behavior -- a very commercial, but no less beneficial, norm. This kind of action is well outside the reach of what most Western governments could achieve, yet it is being done routinely without their needed to be involved.

In future, commercial neutrality will become ever more important as power is likely to continue to shift away from central governments and to non-state actors (like companies). Indeed, could there even be a major cyber conflict if the global network providers (like AT&T, NTT, or BT) decided to suppress it?

8. CONCLUSION

Political neutrality will be an important norm for future cyber conflicts and this paper has examined the idea: what is it, past literature, and important and overlooked aspects. The central part of this paper developed a reasonable, but notional, scenario that explored how various nations would have different levels of neutrality, a determination helped by the ten-point scale of the Spectrum of National Responsibility.

Though the discussion of neutrality in cyber conflict started at least in 1999, with the DoD General Counsel paper, it seems to have made little headway until just the last few years. Further research should extend several of the ideas in this paper, including the difference between political and legal neutrality, the use of the Spectrum of State Responsibility, and include analyses that include the severity, obviousness, stoppability and duration of the attacks in question.

This paper introduced the importance of commercial neutrality, given the outsize role of the private sector in cyberspace. This area deserves much more research, indeed more than is given to exploring how the Hague and other treaties apply.

In future, States and others that see cyber conflicts, like those against Estonia in 2007, are unlikely to be able to sit back and say "not my problem" even as attacks transit their network.

18 William Woodcock, "The Next Fighting Force in Cyberspace," Conference on CyberFutures, Air Force Association, 23 March 2012.

When everyone is a neighbor in cyberspace, there will be no sidelines on which to sit. New norms, some backed with the force of international law, will come into the fore. These and other issues will become increasing important as the world sees more cyber conflicts and the researchers that study and predict it increase our understanding.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Neutrality in Cyberspace

Wolff Heintschel von Heinegg
Faculty of Law
Europa-Universität
Frankfurt (Oder), Germany
heinegg@europa-uni.de

Abstract: The primary object and purpose of the law of neutrality is to protect the (territorial) sovereignty of neutral States and to prevent an escalation of an international armed conflict. Despite of the unique characteristics of cyberspace there is widespread agreement that that body of law applies to cyber operations taken against, or by use of, cyber infrastructure that is located within the territory of neutral States.
Belligerents must respect the inviolability of neutral States and they are prohibited to exercise belligerent rights within their territory. It is, however, not yet sufficiently clear whether that prohibition also applies to (malicious) cyber activities transmitted through neutral cyber infrastructure.
Neutral States are prohibited to allow the exercise of belligerent rights within their territory. Moreover, they are under an obligation to take all feasible measures to terminate such exercise. Again, it is unclear whether neutral States are also obliged to prevent a future exercise of belligerent rights.
If a neutral State is unwilling or unable to comply with its obligation to terminate (or to prevent) a violation of its neutral status, the aggrieved belligerent is entitled to enforce the law of neutrality, subject to proportionality.

Keywords: *neutrality, neutral cyber infrastructure, prohibition of exercising belligerent rights within neutral territory, enforcement of neutral obligations*

1. INTRODUCTION

'Neutrality' denotes the legal status of a State that is not a party to an international armed conflict. Since the rules of international law applicable to neutral States are predominantly laid down in the 1907 Hague Conventions V[1] and XIII[2] one might be inclined to assume that the law of neutrality has become obsolete by desuetude or because an impartial stance vis-à-vis the aggressor and the victim of aggression would be irreconcilable with the *jus ad bellum* as codified in the UN Charter.

Indeed the international armed conflicts that occurred after the end of the Second World War (e.g., the conflicts between Israel and Egypt, India and Pakistan, United Kingdom and Argentina, or Iraq and Iran) might cast doubts on the continuing validity of the traditional law

1 Convention (V) Respecting the Rights and Duties of Neutral Powers and Persons in Case of War on Land, The Hague, 18 October 1907, 2 AJIL Supp. 117-127 (1908).

2 Convention (XIII) Concerning the Rights and Duties of Neutral in Naval War, The Hague, 18 October 1907, 2 AJIL Supp. 202-216 (1908).

of neutrality. It may, however, not be left out of consideration that States, although their conduct may not always have been in compliance with the principle of impartiality, have recognized that the traditional law of neutrality continues to apply to contemporary international armed conflicts.[3] It suffices to refer to the most recent military manuals of the USA[4], Canada[5], the United Kingdom[6] and Germany[7] as well as to the San Remo Manual[8], the ILA Helsinki Principles[9] and the HPCR Manual[10]. Hence, the law of neutrality is well alive.[11]

Under the UN Charter it is, at least in theory, possible to distinguish between an aggressor and the victim of aggression. This, however, does not mean that States are entitled to unilaterally absolve themselves from the obligations of the law of neutrality and to take a 'benevolent' attitude in favour of the alleged victim of an unlawful use of force.[12] If, however, the UN Security Council has decided upon preventive or enforcement measures under Chapter VII of the UN Charter, the scope of applicability of the law of neutrality will be reduced considerably and the 1907 Hague Conventions will be inapplicable.[13] In view of Articles 25 and 103 of the UN Charter States not parties to an international armed conflict are obliged to comply with UN Security Council decisions and in any event to refrain from activities interfering with or impeding the exercise of enforcement operations under such resolution.[14]

Hence, the present paper starts from the premise that, subject to decisions by the UN Security Council under Chapter VII of the UN Charter, the traditional law of neutrality applies to States not parties to an international armed conflict. It will first explore whether and to what extent

3 See Dietrich Schindler, 'Transformations in the Law of Neutrality since 1945', in: *Humanitarian Law of Armed Conflict – Challenges Ahead, Essays in Honour of Frits Kalshoven*, 367-386 (ed. by A.I.M. Delissen/G.J. Tanja, Dordrecht 1991); Wolff Heintschel von Heinegg, 'Wider die Mär vom Tode des Neutralitätsrechts', in: *Crisis Management and Humanitarian Protection, Festschrift für Dieter Fleck*, 221-241 (ed. by H. Fischer et al., Berlin 2004).

4 The Commander's Handbook on the Law of Naval Operations, NWP 1-14M, Chapter 7 (Newport 1997).

5 Law of Armed Conflict at the Operational and Tactical Levels, Chapter 13 (2003).

6 UK Ministry of Defence, *The Manual of the Law of Armed Conflict*, (Oxford 2004). It is important to note that the UK Manual does not contain a chapter specifically devoted to the law of neutrality. However, its continuing validity is expressly recognized in para. 1.42 and Chapters 12 (Air Operations) and 13 (Maritime Warfare) contain rules on neutral States, neutral aircraft and neutral vessels.

7 The Federal Ministry of Defence of the Federal Republic of Germany, Humanitarian Law in Armed Conflicts – Manual, Chapter 11 (Bonn 1992).

8 San Remo Manual on International Law Applicable to Armed Conflicts at Sea, paras. 14 *et seq*., available at: http://www.icrc.org. See also (ed.), San Remo Manual on International Law Applicable to Armed Conflict at Sea (ed. by L. Doswald Beck, Cambridge 1995).

9 ILA, *Helsinki Principles on the Law of Maritime Neutrality*, ILA Report of the Sixty-Eighth Conference, at 497 et seq. (London 1998).

10 Program on Humanitarian Policy and Conflict Research at Harvard University, Manual on International Law Applicable to Air and Missile Warfare, Section X (Bern 2009).

11 See Heintschel von Heinegg (supra note 3), at 232 *et seq*.

12 Wolff Heintschel von Heinegg, ''Benevolent' Third States in International Armed Conflicts: The Myth of the Irrelevance of the Law of Neutrality', in: *International Law and Armed Conflict: Exploring the Faultlines*, 543-568 (ed. by Michael N. Schmitt and Jelena Pejic, Leiden / Boston 2007).

13 See San Remo Manual (supra note 8), paras. 7-9; HPCR Manual (supra note 10), Rule 165; Helsinki Principles (supra note 9), para. 1.2. For the powers of the UN Security Council and the obligations of UN member States see Yoram Dinstein, *War, Aggression and Self-Defence*, at 279 *et seq*., 289 *et seq*. (4th ed., Cambridge 2005). For a restrictive approach to the powers of the UN Security Council see Erika de Wet, *The Chapter VII Powers of the United Nations Security Council*, at 133 *et seq*. (Oxford 2004).

14 For an analysis of the effects of Article 103 UN Charter see Rudolf Bernhardt, 'Article 103', in: The Charter of the United Nations. A Commentary, Vol. II, 1292-1302, at 1295 *et seq*. (ed. by Bruno Simma, 2nd ed., Oxford 2002).

that body of law is applicable to cyberspace (2.) and it will then identify the obligations of belligerents (3.) and neutrals (4.) with regard to (military) operations in cyberspace.

2. APPLICABILITY OF THE LAW OF NEUTRALITY TO CYBERSPACE

The continuing validity of the core principles and rules of the law of neutrality cannot be doubted in the course of an international armed conflict that is characterized by the use of traditional (kinetic) weapons. But when it comes to hostilities and hostile acts conducted in or through cyberspace one might be inclined to reject their applicability. Indeed, if cyberspace is considered to be a new '5th dimension', a 'global common', that "defies measurement in any physical dimension or time space continuum"[15] it could be rather difficult to maintain that the law of neutrality applies. If we acknowledge, however, that cyberspace "requires a physical architecture to exist"[16], many of the difficulties can be overcome.

The law of neutrality serves a double protective purpose. On the one hand, it is to protect the (territorial) sovereignty of neutral States and their nationals against the harmful effects of the ongoing hostilities. On the other hand, it aims at the protection of belligerent interests against any interference by neutral States and their nationals to the benefit of one belligerent and to the detriment of the other. Thus, the rules and principles of the law of neutrality aim at preventing an escalation of an ongoing international armed conflict "in regulating the conduct of belligerents with respect to nations not participating in the conflict, in regulating the conduct of neutrals with respect to belligerents, and in reducing the harmful effects of such hostilities on international commerce."[17]

Applied to the cyber context it is safe to conclude that the law of neutrality protects the cyber infrastructure that is located within the territory of a neutral State or that profits from the sovereign immunity of platforms and other objects used by the neutral State for non-commercial government purposes.[18] Hence, belligerents are under an obligation to respect the sovereignty and inviolability of States not parties to the international armed conflict by refraining from any harmful interference with the cyber infrastructure located within neutral territory. Neutral States must remain impartial and they may not engage in cyber activities that support the military action of one belligerent and that are to the detriment of the other belligerent. Moreover, they are obliged to take all feasible measures to terminate an abuse of the cyber infrastructure located within their territory (or on their sovereign immune platforms) by any of the belligerents.

The correctness of these findings might be doubted because they are based upon a teleological interpretation of the law of neutrality alone. However, they are supported not only by the

15 Thomas Wingfield, *The Law of Information Conflict: National Security Law in Cyberspace*, at 17 (Aegis Research Corp. 2000).

16 Patrick W. Franzese, 'Sovereignty in Cyberspace: Can It Exist?', 64 AFLR 1-42, at 33 (2009).

17 NWP 1-14M (*supra* note 4), para. 7.1.

18 Territory consists of the land territory, the internal waters, the territorial sea and, where applicable, the archipelagic waters of a neutral State as well as the airspace above those areas. Platforms and objects enjoying sovereign immunity include warships, military aircraft and diplomatic premises and communication devices.

majority of authors dealing with the issue of neutrality in the cyber context[19] but also by State practice. For instance, the U.S. Department of Defense (DoD) has taken the position that "long-standing international norms guiding state behavior – in times of peace and conflict – also apply in cyberspace."[20] The DoD Cyberspace Policy Report, *inter alia*, emphasizes that "applying the tenets of the law of armed conflict are critical".[21] The Report also addresses activities "taking place on or through computers or other infrastructure located in a neutral third country".[22] It may be added in this context that the applicability of the law of neutrality to cyberspace has also been acknowledged in the recent HPCR Manual.[23] Since that Manual has been endorsed by a considerable number of governments it may be considered as a restatement of the existing law and as reflecting the consensus of States on the issues dealt with in the Manual.

Of course, the rules of the traditional law of neutrality, while in principle applicable to cyberspace, may require clarifications or even modifications because of the unique characteristics of cyberspace.[24] Still, the "law of armed conflict and customary international law […] provide a strong basis to apply such norms to cyberspace governing responsible state behavior."[25]

3. OBLIGATIONS OF BELLIGERENTS

According to the law of neutrality belligerents are obliged to respect the inviolability of neutral territory. Hence, they are prohibited to conduct hostilities, to exercise belligerent rights or to establish bases of operations within neutral territory. These prohibitions are laid down in international treaties[26] and they are considered as customary in character.[27]

A. No Harmful Interference with Neutral Cyber Infrastructure

It follows from the foregoing that the cyber infrastructure located within the territory of a neutral State is protected against any harmful interference by the belligerents. It does not matter whether the respective cyber infrastructure is owned (or exclusively used) by the government, by corporations or by private individuals. Neither does the protection depend upon the nationality of the owner. In view of the principle of sovereign immunity the same

19 See, *inter alia*, Joshua E. Kastenberg, 'Non-Intervention and Neutrality in Cyberspace: An Emerging Principle in the National Practice of International Law', 64 AFLR 43-64, at 56 et seq. (2009); Graham H. Todd, 'Armed Attack in Cyberspace: Deterring Asymmetric Warfare with an Asymmetric Definition', *ibid*. 65-102, at 90 *et seq*.; George K. Walker, 'Information Warfare and Neutrality', 33 Vanderbilt J.Trans.L., 1079-1202, at 1182 *et seq*.

20 U.S Department of Defense, *Strategy for Operating in Cyberspace*, at 9 (available at: http://www.defense.gov/news/d20110714cyber.pdf).

21 U.S. Department of Defense, *Cyberspace Policy Report - A Report to Congress Pursuant to the National Defense Authorization Act for Fiscal Year 2011*, Section 934, at 7 et seq. (November 2011), available at: http://www.defense.gov/home/features/2011/0411_cyberstrategy/docs/NDAA%20Section%20934%20Report_For%20webpage.pdf.

22 *Ibid*., at 8.

23 *Supra* note 10, Rule 168 (b).

24 Cyber Policy Report (supra note 21), at 7.

25 *Ibid*., at 8.

26 Articles 1, 2, and 3 of the 1907 Hague Convention V (supra note 1); Articles 1, 2 and 5 of the 1907 Hague Convention XIII (*supra* note 2).

27 See NWP 1-14M (*supra* note 4), para. 7.3; German Manual (*supra* note 7), paras. 1108, 1149; San Remo Manual (*supra* note 8), para. 15; HPCR Manual (*supra* note 10), Rule 166. See also Articles 39, 40, 42 and 47 of the Rules of Aerial Warfare, The Hague, 1923, 32 AJIL Suppl. 12-56 (1938).

protection applies to every cyber infrastructure located on neutral state ships and state aircraft or in diplomatic premises.

The prohibition of harmfully interfering with neutral cyber infrastructure is not limited to cyber attacks, i.e., to cyber operations that cause, or are expected to cause, damage, destruction, death or injury. Rather, it is to be understood as also comprising all activities, whether kinetic or cyber, that either have a negative impact on the functionality or make their use impossible. In other words, it is prohibited to engage in "the use of network-based capabilities […] to disrupt, deny, degrade, manipulate, or destroy information resident in computers and computer networks, or the computers and networks themselves"[28], of a neutral State.

Of course, mere intrusion into neutral cyber infrastructure is not covered by this prohibition because international law lacks a prohibition of espionage. It must be borne in mind that the principle of territorial sovereignty includes the prohibition of exercising jurisdiction on foreign territory.[29] Hence, a cyber operation that may be characterised as an exercise of jurisdiction would be in violation of the sovereignty of the target State. However, that prohibition is of a general character and thus not part of the law of neutrality *strictu sensu*.

B. Exercise of Belligerent Rights and Use of Belligerent Cyber Infrastructure

Belligerents are prohibited to use neutral cyber infrastructure for the purpose of exercising belligerent rights against the enemy or against others. It is important to note that the term 'belligerent rights' is not limited to (cyber) attacks but that it refers to all measures a belligerent is entitled to take under the law of armed conflict against the enemy belligerent, enemy nationals or the nationals of neutral States.[30] This prohibition follows from the very object and purpose of the law of neutrality, i.e., to prevent an escalation of the international armed conflict.

In view of its object and purpose this prohibition also applies to the exercise of belligerent rights by the use of neutral cyber infrastructure that enjoys sovereign immunity because it is used by the organs of a neutral State for exclusively non-commercial government purposes and that is located outside neutral territory. It is not equally clear whether the prohibition also applies to the use (or: abuse) of cyber infrastructure located outside neutral territory that is owned by a private corporation or individual. Be that as it may. In such a situation the respective cyber infrastructure may be considered as contributing to the enemy's military action and the opposing belligerent would therefore be entitled to treat it as a lawful military objective.[31]

Moreover, a belligerent may not make use of its cyber infrastructure for military purposes if it is located on neutral territory. It is irrelevant whether the cyber infrastructure has been 'erected'

28 Arie J. Schaap, 'Cyber Warfare Operations: Development and Use under International Law', 64 AFLR, 121-173, at 127 (2009).

29 Permanent Court of International Justice, Judgment No. 9, *The Case of the S.S. "Lotus"*, PCIJ Ser. A No. 10 (1927), at 18: "La limitation primordiale qu'impose le droit international à l'Etat est celle d'exclure – sauf l'existence d'une règle permissive contraire – tout exercice de sa puissance sur le territoire d'un autre Etat. Dans ce sens, la juridiction est certainement territoriale; elle ne pourrait être exercée hors du territoire, sinon en vertu d'une règle permissive découlant du droit international coutumier ou d'une convention."

30 Such actions comprise detention, requisitions, capture and interception.

31 For the definition of lawful military objectives see Article 52 (2) of the 1977 Additional Protocol I to the 1949 Geneva Conventions. This definition reflects customary international law.

prior to or after the outbreak of the international armed conflict. This prohibition follows from Article 3 of the 1907 Hague Convention V according to which "belligerents are […] forbidden to:

(a) Erect on the territory of a neutral Power a wireless telegraphy station or other apparatus for the purpose of communicating with belligerent forces on land or sea;
(b) Use any installation of this kind established by them before the war on the territory of a neutral Power for purely military purposes, and which has not been opened for the purpose of public messages."

C. Exceptions to the Prohibition of Exercising Belligerent Rights?

As already mentioned, the prohibition of exercising belligerent rights by the use of neutral cyber infrastructure must be interpreted in the light of the unique characteristics of cyberspace.[32] Cyberspace is an "interdependent network of information technology infrastructures, including the Internet, telecommunications networks, computer systems, and embedded processors and controllers".[33] In view of the interdependence and the ubiquity of cyberspace and its components it would be almost impossible for a belligerent to prevent the routing of malicious data packages through the cyber infrastructure located within the territory of a neutral State although it is ultimately aimed against the enemy.

Therefore it seems to be logical and perhaps even cogent to apply Article 8 of the 1907 Hague Convention V to cyber operations and to cyber attacks conducted by a belligerent against its enemy. Article 8 provides:

> "A neutral Power is not called upon to forbid or restrict the use on behalf of the belligerents of telegraph or telephone cables or of wireless telegraphy apparatus belonging to it or to companies or private individuals."

Although some doubts have been articulated in the literature as to whether Article 8 Hague V was at all applicable to cyberspace[34], that position would not justify a total rejection of Article 8 because it is based on the assumption that a cyber operation conducted through neutral cyber infrastructure is to be considered as originating from neutral territory. Then, and only then, it would have to be considered an exercise of belligerent rights from neutral territory.

It must be borne in mind, however, that Article 8 only applies to communications and that Article 2 of Hague Convention V prohibits belligerents, *inter alia*, to "move […] munitions of war or supplies across the territory of a neutral Power". If the distinction between mere communications and a passage of "munitions of war" were applied to cyberspace any transmission of a 'cyber weapon' through neutral cyber infrastructure would constitute a violation of the law of neutrality. Indeed, there are some indications that States will share that view. For instance, the

32 *Supra* note 24 and accompanying text.
33 Joint Chiefs of Staff, Joint Pub. 1-02, Dept. of Defense Dictionary of Military and Associated Terms, at 41 (12 April 2001). See also the definition by Schaap, *supra* note 28, at 126, who defines 'cyberspace' as a "domain characterized by the use of [computers and other electronic devices] to store, modify, and exchange data via networked systems and associated physical infrastructures".
34 Kastenberg (*supra* note 19), at 56 et seq.; Todd (*supra* note 19), at 90 et seq.

Office of General Counsel of the U.S. DoD, in 1999, arrived at the conclusion that "[t]here is nothing in this agreement [i.e., Hague Convention V] that would suggest that it applies to systems that generate information, rather than merely relay communications."[35] It is interesting to note that the U.S. DoD seems to be prepared to apply Article 8 of the 1907 Hague Convention V to cyberspace, although it would limit its applicability to mere communications, i.e., to cyber operations that do not amount to a cyber attack.

It may, however, not be left out of consideration that Articles 2 and 8 of the 1907 Hague Convention V are based on the assumption that a neutral State exercises full and effective control over its entire territory but not over installations and objects used for communications purposes. The different degrees of feasible and effective control must also be taken account of in the cyber context. This especially holds true for a "public, internationally and openly accessible network such as the Internet". Hence, the HPCR Manual provides:

> "[W]hen Belligerent Parties use for military purposes a public, internationally and openly accessible network such as the Internet, the fact that part of this infrastructure is situated within the jurisdiction of a Neutral does not constitute a violation of neutrality."[36]

It must be noted that the HPCR Manual does not distinguish between mere communications on the one hand and the transmissions of cyber weapons on the other hand. The phrase "use for military purposes" is sufficiently broad to cover both. This seems to be a reasonable adaptation of the traditional rules of the law of neutrality to cyberspace. Because of the complexity and interdependence of contemporary networks, such as the Internet, it is impossible to effectively exercise the control necessary for an effective interference with communications over such networks. This is underlined by the fact that most such communications are often neither traceable nor predictable since they will be transmitted over lines of communications and routers passing through various countries before reaching their ultimate destination. Therefore, the mere fact that military communications, including cyber attacks, have been transmitted via the cyber infrastructure of a neutral State might not be considered a violation of that State's neutral obligations.

It is admitted that despite of the attractiveness of the HPCR Manual's approach for both belligerents and neutral States it is far from clear whether such a far-reaching adaptation of Article 8 Hague V to cyber operations conducted for military purposes will ultimately be accepted as reflecting contemporary customary international law. Modern State practice, especially the cyber operations during the 1999 Kosovo Campaign, the conflicts in Afghanistan (2001) and Iraq (2003), and the armed conflict between Georgia and Russia (2007), does not necessarily provide sufficient evidence that any cyber operation, including the transmission of cyber weapons, through neutral cyber infrastructure does not constitute a violation of neutrality. On the one hand, there is no unclassified information that the respective cyber operations did amount to cyber attacks and that they had been routed through neutral cyber infrastructure. The DDoS attacks against Georgia, according to the position taken here, do not qualify as cyber attacks and can therefore not be assimilated to the transit of "munitions of war" under Article 2 of Hague Convention V. On the other hand, the U.S. DoD Cyberspace Policy Report seems

35 U.S. Department of Defense, Office of General Counsel, An Assessment of International Legal Issues in Information Operations, at 10 (Washington, D.C., May 1999), available at: http://www.au.af.mil/au/awc/awcgate/dod-io-legal/dod-io-legal.pdf.

36 HPCR Manual (*supra* note 10), Rule 167(b).

to justify the conclusion that the U.S. government is prepared to consider every "malicious cyber activity" as in violation of the law of neutrality irrespective of whether they have been launched from or merely transmitted through "computers or other infrastructure located in an neutral third country".[37]

Hence, it may be held that the use of neutral cyber communications by a belligerent does not constitute a violation of neutrality even though it serves military purposes. However, it is less clear whether this finding also holds true if the cyber operation in question qualifies as a 'malicious cyber activity' or as a cyber attack. We will return to this issue in the context of the consequences of a violation of the law of neutrality by neutral States.

4. OBLIGATIONS OF NEUTRAL STATES

The law of neutrality, in view of its object and purpose[38], poses obligations not only upon the belligerents but also on neutral States. Leaving aside the duty of impartiality[39], these obligations may be divided into three categories: (1) prohibition to allow or to tolerate the exercise of belligerent rights; (2) obligation to terminate (and probably to prevent) a violation of neutrality by a belligerent; and (3) obligation to tolerate the enforcement of the law of neutrality by the aggrieved belligerent.

A. Prohibition of Tolerating the Exercise of Belligerent Rights

According to Article 5 of the 1907 Hague Convention V a "neutral Power must not allow any of the acts referred to in Articles 2 to 4 to occur in its territory". Accordingly, a neutral State is prohibited to allow or to tolerate the exercise of belligerent rights from the cyber infrastructure located within its territory or that is located outside its territory, provided that the neutral State exercises exclusive control over it.[40]

It may be noted that the different interpretations of Article 8 of Hague V may have far-reaching consequences. According to the approach taken in the HPCR Manual[41], a malicious cyber activity routed through neutral cyber infrastructure that is a component of, e.g., the Internet, would not constitute a prohibited exercise of belligerent rights. Hence, a neutral State allowing or tolerating such an activity would not violate its obligations under the law of neutrality. If the HPCR approach is not considered as reflecting customary international law, the transmission of a cyber attack through neutral infrastructure would have to be considered a prohibited exercise of belligerent rights and the neutral State allowing or tolerating the transmission would be in violation of its neutral obligations.

But even if the latter approach is taken the consequences are less grave than one may assume.

37 *Supra* note 21, at 8.
38 *Supra* note 17 and accompanying text.
39 Article 9 of the 1907 Hague Convention and Article 9 of the 1907 Hague Convention XIII provide that "every measure of restriction or prohibition taken by a neutral Power […] must be impartially applied by it to both belligerents." Hence, restrictions on military communications via its cyber infrastructure must be applied impartially by the neutral State. See also San Remo Manual (*supra* note 8), para. 19.
40 *Supra* 3. A.
41 *Supra* note 36.

Contrary to a position taken in the literature[42], the use of the term "allow" in the traditional rule presupposes knowledge by the organs of the neutral State. That will be the case if the organs have detected a malicious cyber activity/cyber attack or if they have been informed, e.g. by the other belligerent and in a sufficiently credible manner, that the activity has originated from, or has been transmitted through, the respective neutral State's cyber infrastructure. However, such knowledge will result in a violation of the law of neutrality by the neutral State only if the malicious cyber activity continues. In most cases, cyber attacks will occur at a considerably high speed so that *ex-post-facto* knowledge can hardly suffice to justify a claim of a violation of the law of neutrality. And even if one were prepared to consider constructive (as opposed to actual) knowledge as sufficient for a violation of the said obligation that would not result in noticeable changes. Constructive knowledge means that the organs of a neutral State should have known of the malicious activity. Again, in most cases constructive knowledge would not necessarily result in a violation of neutral obligations.

This would probably be different if, as a result of the prohibition of allowing the exercise of belligerent rights, neutral States were obliged to actively monitor cyber activities originating from or transiting through their cyber infrastructure. However, it is far from settled whether such an obligation exists. Of course, the San Remo Manual, *inter alia*, provides that a "neutral State must take such measures […], including the exercise of surveillance, as the means at its disposal allow, to prevent the violation of its neutrality by belligerent forces."[43] It is, however, not likely that especially those States that defend the freedom of Internet communications will agree that the obligation to monitor territory and certain sea areas applies equally to the cyber infrastructure located in their territory.

B. Obligation to Terminate (and to Prevent) a Violation of Neutrality

According to the traditional law of neutrality, neutral States are obliged to terminate an exercise of belligerent rights and any other violation of its neutrality by one of the belligerents.[44] This obligation is part of contemporary customary international law.[45]

The obligation to enforce its neutral status against violations by the belligerents is not absolute in character but it is limited to what is feasible. In other words, the neutral State is obliged to use all means reasonably available to it to terminate an exercise of belligerent rights within its territory.[46] The applicable standard is, thus, not objective but rather subjective. Everything will depend on the means and capabilities factually available to the respective neutral State. It needs to be emphasized that, subject to feasibility, the duty to enforce its neutral status entails an obligation to use all means necessary to effectively terminate an unlawful exercise of belligerent rights. This may include the use of force. The belligerent against whom such enforcement measures are applied may not consider them as a hostile act, i.e., it is obliged to tolerate them.[47]

42 Kastenberg (*supra* note 19), at 57.
43 San Remo Manual (*supra* note 8), para. 15.
44 *Ibid*., paras. 18 and 22; HPCR Manual (*supra* note 10), Rule 168(a). See also Articles 42 and 47 of the 1923 Hague Rules (*supra* note 27).
45 San Remo Manual (*supra* note 8), para. 22; HPCR Manual (*supra* note 10), Rule 168(a); NWP 1-14M (*supra* note 4), para. 7.3; German Manual (*supra* note 7), para. 1109.
46 *Ibid*.
47 Article 10 of the 1907 Hague Convention V; HPCR Manual (supra note 10), Rule 169; Hague Rules (*supra* note 27), Article 48.

The obligation to terminate an ongoing (!) violation of neutrality presupposes – actual or constructive – knowledge on part of the organs of the neutral State.[48] It is quite probable that the neutral State is unaware of an abuse of its cyber infrastructure. But even if such actual or constructive knowledge existed it would in most cases be futile to demand from the neutral State to take measures against the respective belligerent because the cyber operation triggering the duty to terminate will no longer continue.

Obviously, such a limitation to ongoing (malicious) cyber activities is considered by some authors to be insufficient. They therefore claim that a neutral State is also obliged to take all feasible measures to prevent an exercise of belligerent rights, i.e., before it occurs.[49] At first glance, that position seems to reflect customary international law because some military manuals expressly refer not only to an obligation to terminate an ongoing violation of neutrality but also to a duty to prevent an exercise of belligerent rights within neutral territory.[50] It is, however, doubtful, whether the use of the term "prevent" is meant to establish an obligation vis-à-vis future violations of neutrality. But even if that were the case, the duty to prevent would be limited to territory and national airspace. It is far from clear whether States are willing to accept it when it comes to the use of their cyber infrastructure because that would imply an obligation to continuously monitor cyber activities originating from or transiting through their cyber infrastructure. Moreover, the identification of the malicious nature of data packages transiting through a network would in most cases be most difficult, if not impossible.

Therefore, there are good reasons for rejecting a (prospective) duty of prevention. If at all, such an obligation would only exist with regard to activities within neutral territory that could be assimilated to those covered by Article 8 of the 1907 Hague Convention XIII.[51] For instance, the authorities of a neutral State may have (actual or constructive) knowledge of the activities of a group of hackers that has been employed by a belligerent government to develop a cyber weapon that is to be used against the enemy. In such a situation the neutral State would be obliged to take all feasible measure to prevent the departure of the cyber weapon from its territory (jurisdiction).

C. Consequences of Non-Compliance by Neutral States

Admittedly, during the international armed conflicts since the end of the Second World War neutral States have regularly not complied with their obligations under the law of neutrality.[52] They either openly or clandestinely assisted one party to an international armed conflict to the detriment of the other belligerent. However, already the fact that some neutral governments have tried to conceal their 'unneutral service' is sufficient evidence that they considered themselves bound by the law of neutrality. And even those governments that openly supported one side of an international armed conflict took pains in justifying their conduct. Eventually they were in

48 *Supra* 4. A.

49 Kastenberg (*supra* note 19), at 56 *et seq*.

50 San Remo Manual (*supra* note 8), para. 15; HPCR Manual (*supra* note 10), Rule 168(a); NWP 1-14M (*supra* note 4), para. 7.3.

51 "A neutral Government is bound to employ the means at its disposal to prevent the fitting out or arming of any vessel within its jurisdiction which it has reason to believe is intended to cruise, or engage in hostile operations, against a Power with which that Government is at peace. It is also bound to display the same vigilance to prevent the departure from its jurisdiction of any vessel intended to cruise, or engage in hostile operations, which had adapted entirely or partly within the said jurisdiction for use in war."

52 See Heintschel von Heinegg (*supra* note 12), at 556 *et seq*.

a comfortable position in view of the fact that the aggrieved belligerent was unable to react to their non-compliance with neutral obligations.

The law of neutrality provides that if a neutral State fails to terminate (or prevent) an exercise of belligerent rights or another violation of neutrality by one belligerent, the other belligerent is entitled to take the measures necessary to terminate the violation.[53] The right of the aggrieved belligerent to enforce the law of neutrality comes into operation if the neutral State is either unwilling or unable to comply with its obligation to terminate (or prevent) a violation of its neutral status by the enemy. This right is a specific form of a counter-measure, i.e., a measure that would be unlawful were it not taken in response to a violation of international obligations by the target State.[54] Its object and purpose is (1) to induce the neutral State to comply with its obligations; and (2) to enable the aggrieved belligerent to preserve its security interests. Hence, not every violation of the neutral status by one belligerent justifies a resort to counter-measures by the other belligerent. The violation in question must have a negative impact on the legitimate security interests of that belligerent. This will not be the case if a belligerent takes measures against a neutral State's cyber infrastructure that do not imply a military advantage over the enemy. The right to respond to the violation is then exclusively reserved to the neutral State. Moreover, the exercise of the right is probably subject to a *de minimis* exception.

Moreover, the aggrieved belligerent is not entitled to immediately resort to the exercise of counter-measures. For instance, the San Remo Manual provides: "If the neutral State fails to terminate the violation of its neutral waters by a belligerent, the opposing belligerent must so notify the neutral State and give that neutral State a reasonable time to terminate the violation by the belligerent."[55] An immediate response by the aggrieved belligerent is lawful only, if

- the violation constitutes a serious and immediate threat to the security of that belligerent;
- there is no feasible and timely alternative; and
- the enforcement measure taken is strictly necessary to respond to the threat posed by the violation.[56]

The aggrieved belligerent's right to enforce the law of neutrality certainly applies to cyberspace if a malicious cyber activity originates from within the territory of a neutral State.[57] The U.S. DoD seems to be prepared to take such enforcement measures if it is possible to determine that a neutral State is aware of a malicious cyber activity within neutral territory. The DoD will take account of the following aspects:

- "The nature of the malicious cyber activity;
- The role, if any, of the third country;
- The ability and willingness of the third country to respond effectively to the malicious cyber activity; and

53 NWP 1-14M (*supra* note 4), para. 7.3; San Remo Manual (*supra* note 8), para. 22; HPCR Manual (*supra* note 10), Rule 168(b).
54 International Law Commission, *Responsibility of States for Internationally Wrongful Acts*, Articles 22, 49-54, U.N. Doc. A/56/10.
55 San Remo Manual (*supra* note 8), para. 22.
56 *Ibid*. See also HPCR Manual (*supra* note 10), Rule 168(b).
57 See the Cyberspace Policy Report (*supra* note 21), at 8.

- The appropriate course of action for the U.S. Government to address potential issues of third-party sovereignty depending upon the particular circumstance."[58]

This is a clear restatement of the rules of the law of neutrality and it gives sufficient evidence of the DoD's willingness to apply those rules to conduct in cyberspace.

5. CONCLUSIONS

It has been shown that the traditional law of neutrality is, in principle, applicable to cyberspace, especially to belligerent cyber operations that violate the status of neutral States because they qualify as an exercise of belligerent rights within neutral territory. The special characteristics of cyberspace do not as such pose an obstacle to such application. However, there certainly remains an urgent need for clarification and even adaptation of the traditional law. In view of the interdependence of the networks through which data are transmitted and their potentially disastrous effects on critical infrastructure there is a high probability that belligerent States will take measures against neutral States and their respective cyber infrastructure, including the use of (kinetic) force if they must assume that vital security interests are at stake. Such measures have the potential of jeopardizing the essential object and purpose of the law of neutrality, i.e., preventing an escalation of an international armed conflict. The U.S. government has taken first steps that are most helpful in the identification of the applicable rules of international law and their interpretation in the light of the challenges brought about by the specific characteristics of cyberspace. The U.S. government should continue those efforts and other governments should closely cooperate with the U.S. government with a view to arriving at an operable consensus that takes into consideration global interoperability, network stability, reliable access and cyber security due diligence.[59]

58 *Ibid*.
59 U.S. President, *International Strategy for Cyberspace*, at 10 (May 2011).

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Impact of Cyberspace on Human Rights and Democracy

Vittorio Fanchiotti
Faculty of Law
University of Genova
Genova, Italy
vittorio@unige.it

Jean Paul Pierini
Fleet Command
Italian Navy
Rome, Italy
pierini.jeanpaul@libero.it

Abstract: This paper focuses on the asserted 'boundlessness' of cyberspace in order to examine how and to what extent jurisdiction, in its various meaning and forms (jurisdiction to prescribe, to adjudicate and to execute), over activities taking place in the cyberspace may be asserted and even exercised, based on traditional jurisdictional links and also on new trends. The paper also examines conflicts of law in civilian (mainly tort laws and laws on the protection of rights of the personality as well as intellectual property) and criminal matters. Determining what set of rules applies to a certain fact or situation implies a reference to those rules establishing where such a fact or situation has legally taken place and is to be localised (*locus commissi delicti*), and a reference to main criteria including those focusing on the conduct, the localisation of the hardware, the effect, the access to the informatics system, the accessibility of the information and future trends. The paper further highlights that the enforcement of activities in cyberspace appears to be affected by an assimilation to traditional forms of investigative activities, such as search or inspection or even the interception of communication or data flow, which are to a certain degree misleading in respect of the specific means employed. A specific reference to the role of providers in enforcement activities is also included. The second part of the paper deals with the traditional human rights relevant to cyberspace and to the broader concept of 'right to access' cyberspace, as well as the uncertainties derived from the fact that a plurality of state and non-state actors may limit and interfere with human rights in cyberspace. The paper specifically deals with the commercial dimension of cyberspace and with eventual corporate liability for human rights violation (multinational corporations violating rights to privacy in connection with or on behalf of states or enforcing censorship) based on US legislation and also taking into consideration European trends. The paper finally highlights the supportive role to the protection of human rights of regulatory bodies enforcing fair-trade and anti-trust regulations, and the multinational dimension of free trade in promoting human rights, by eventually considering restrictions in cyberspace and censorship as restrictions to trade under WTO agreements.

Keywords: *jurisdiction, enforcement, conduct, effects, antitrust, fair-trade, censorship, WTO*

1. IS CYBERSPACE REALLY WITHOUT BOUNDARIES?

Cyberspace, which is still lacking a standard and universally accepted definition, identifies a domain encompassing the digitalised information itself, as well as the infrastructure (including satellite telecommunications), server networks, computers and especially the internet, that makes the spectrum useful. However, cyberspace is mostly defined by how it is used and is identified with the World Wide Web.[1]

Accessing and transmitting information through the World Wide Web has become a significant part of contemporary lifestyles and entertainment, and has progressively developed into an awareness of a global community where the individual has the ability to connect socially and directly with other individuals without apparent political, social or racial borders. So called 'second life' social experiences, where the individual shows up through an 'avatar' giving them their identity of choice, have also reinforced the idea of cyberspace as a domain in which the individual may find, develop and exploit their own 'parallel reality'.

Influential literature from almost four decades ago significantly altered the perception of cyberspace and the web and, together with an increased awareness of the right to access directly information and knowledge (also as a substitute for declared but not sufficiently implemented human and social rights, to include the right to information), encouraged the perception of cyberspace as a 'global common'. The latter concept encompasses those goods and rights which are not suitable for appropriation by any state, entity or individual.

While 'virtual reality' has heavily contributed to the misconception of cyberspace as a space not marked or flagged by any state sovereignty, control or even governance, the potential for social connection and direct access to information and knowledge has contributed to the idea that within cyberspace (and specifically for those accessing it) exchange of information should be free and unhampered by rules and laws.

This said, in the authors' views, the assessment of the legal consequences of phenomena taking place in cyberspace, and the evaluation of the consequences of setting roles, should not be affected by the suggestion of cyberspace as a non-physical realm and, in general terms, by cyberpunk literature, as such phenomena always have a specific physical dimension. They are also linked to a clear geographical dimension represented by server location, point of access, human conduct and a legally appreciable effect.[2] Accordingly, the legal reasoning should not be altered by the social perception of cyberspace which is, to a substantial part, influenced by

1 For this purpose see Maj. Gen. Mark Barret, Dick Bedford, Elizabeth Skinner & Eva Vergles, *Assured Access to the Gobal Commons*, Supreme Allied Command Transformation, North Atlantic Treaty Organization, Norfolk Virginia, USA, April 2011, p. 35.

2 The idea is nonetheless reflected in the Explanatory Report to the European Cybercrime Convention, adopted on the 8th of November, 2001 at Budapest, CETS/SEV No 185, recalling, at § 7, the decision CDPC/103/211196, of the European Committee on Crime Problems (CDPC) reached in November 1996, stating that 'by connecting to communication and information services users create a kind of common space, called 'cyber-space', which is used for legitimate purposes but may also be the subject of misuse. These 'cyber-space offences' are either committed against the integrity, availability, and confidentiality of computer systems and telecommunication networks or they consist of the use of such networks of their services to commit traditional offences. The transborder character of such offences, e.g. when committed through the Internet, is in conflict with the territoriality of national law enforcement authorities'.

the idea that the web pages visited are hosted on a server somewhere in the unknown, simply because it is more exciting than associating the IP address to a certain location.

Wrongs committed through information technology almost entirely rely on the transmission of information and very often displace law enforcement mechanisms which do not benefit from equally advanced forms of judicial and police cooperation, effectively having the effect of a 'force multiplayer'. As such, certain legal difficulties in the repression of crimes, and also civilian torts, may have also contributed to the perception of cyberspace as a borderless domain, overarching a context of legal systems, each of which is jealous of its own prerogatives. Clearly, those willing to commit crimes find it easier to cross borders through the web than a law enforcement officer does in order to deter, stop, take evidence or arrest.

On the other hand, 'the critical nodes, or "gateways" to cyberspace … are entirely in the hands of commercial enterprises … internet service providers (ISPs) connect computers to the internet, while web hosting services maintain websites on the World wide web … browsers like Internet Explorer, Safari, Chrome, and Firefox make such content accessible.'[3]

Significant interest has been raised by 'cyber attacks' from the military perspective as an autonomous pillar standing aside from the traditional fight against cyber crime, with which it shares uncertain borders. This may be considered a consequence of the yet to be clearly determined threshold at which criminal activity becomes a military attack which may trigger the use of force in self-defence (to include offensive cyber responses under traditional principles of international humanitarian law, once the crucial issue of distinction between civilians and combatants/civilians directly taking part in hostilities has been addressed in a satisfactory way), but is also due to the fact that the reaction to a direct armed attack is easier to justify than violating the sovereignty of another state for a cross-border arrest.

Not being influenced by an indeterminist notion of one or more 'parallel virtual' universes does not mean that legal concepts, especially those defining enforcement activities (e.g. online searches) should not evolve to take technical developments into consideration and be correspondingly adapted in order to be more effective and also to preserve the essence of guarantees.

The social perception of cyberspace, and furthermore the role played by cyberspace in making knowledge available and a global community accessible, should be clearly taken into consideration when it comes to ascertaining if democratic rights stated in modern constitutions and human rights instruments have changed their essence and now encompass, through the web, a new dimension. These rights include those dealing with freedom of speech, the right to express opinions, the right to information (as well as the right to inform), and associative rights, as well as the right of the individual to develop him/herself in a social context.

On the other side of the coin, depicting the access to cyberspace is the dimension of data protection and the right to informational self-determination of the individual; a right which is endangered by the delocalisation of addresses.

3 Maj. Gen. Mark Barret, Dick Bedford, Elizabeth Skinner, Eva Vergles, *Assured Access to the Gobal Commons*, Supreme Allied Command Transformation, North Atlantic Treaty Organization, Norfolk Virginia, USA, April 2011, p. 35.

This new dimension of the right to access knowledge may pose specific problems with respect to copyright and related rights, which may in the future encounter the same limits as those set to advances in biotechnology and certain patents, and specifically the avoidance of 'excessive protection'.

2. CYBERSPACE AND STATE JURISDICTION

Despite the deceptive reference to cyberspace as a domain without boundaries, phenomena taking place within such domains are based on several criteria subject to state jurisdiction to prescribe, to adjudicate and to execute.

Establishing where a certain activity fulfilling the conditions for the application of a criminal provision took place has always raised legally harsh questions.

'Jurisdiction to prescribe', which identifies with the ambit of application of substantive laws which are eminently territorial, encounters in general terms the main limit of the prohibition to interfere with domestic issues of another state. Prescriptions issued extraterritorially to nationals and foreigners may be subject to the double criminality requirement as a postulate of justice. This requirement is commonly waived when the prescription pertains to the protection of core interests of the state or interests whose protection is generally recognised. In both cases, the effective protection of the interest may not be conditioned by the attitude of the state on whose territory the conduct took place.

Cyberspace is a highly regulated domain in which the territorial regulations may define legal obligations of ISPs, web hosting, commercial enterprises relying on the web in order to do their business, search motors, hardware, software and application producers and sellers, internet points, hot-spots, those collecting, storing, analysing and transmitting personal data, individuals accessing the web and downloading data, and even those travelling through the territory of the state with devices containing stored information. Territorial prohibitions under criminal law may well pertain to the establishment of criminal sanctions for hacking and illegal access to information systems located within the territory or accessed from such a territory, or simply disrupting public services on the territory of such a state.

Regulations eventually perceived as 'extraterritorial' may pertain to content of internet pages accessible from the concerned state, whereas evidence of access may well suit the requirements for the territorial commission of certain crimes, as in the case of libel, racist and xenophobic material, denial, gross minimisation, approval or justification of genocide, or crimes against humanity.[4]

Despite the instant or almost instant character of data transmission, consequences of the transit of certain information through the infrastructure and nets located in the territory of a certain state could be considered by the state for the exercise of jurisdiction to prescribe, based on the existence of available technical means. Current rules developed under international law on jurisdiction over space objects may sustain the exercise of jurisdiction to prescribe in respect of data flow through communication satellites.

4 For this purpose see the 'Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems', produced in Strasbourg on 28 January 2003, CETS/SEV No 189.

One current trend in law-making may be seen in the establishment of obligations for commercial enterprises offering cyberspace-related issues which are invested by a 'guarantee position' and increasingly held liable, either criminally and/or under tort laws, for content of hosted web pages including child pornography, commercial offers of counterfeit goods, libellous content and violation of copyright and related rights.[5] Such enterprises are encouraged to develop commercial arrangements with those benefitting from their services and establish proper procedures in order to control such contents through technical means.

The debate on the relevance of the conduct or the effect caused by such conduct, in order to establish the jurisdiction to adjudicate as a consequence of the definition of the crime as 'territorial' under the criteria for the establishment of the *locus commissi delicti*, dates back more than a century to the so-called 'Cutting case' of 1886.[6] This concerned the publication by a US citizen in the border town of El Paso of a defamatory article against a Mexican citizen, where the newspaper circulated in the Mexican city of Paso del Norte. Currently there is a wide practice for the sufficiency of either the conduct, or the realisation of part of the effect the criminal provision was aimed at preventing, in the territory of a certain state in order to consider the crime committed in the territory of that state (so-called ubiquity theory).

A recent issue which falls in between the jurisdiction to prescribe and the jurisdiction to adjudicate (at least where the latter is a consequence of the way the incriminating provision is drafted) is represented by the disclosure through the posting through an access point in State A of information classified in State B.

For this purpose it should be observed that some states consider a crime to have been committed in their territory when the crime aimed to realise its effects there but did not do so or, with respect to the crime of conspiracy, the (foreign) conspiracy aimed to commit a crime in the territory of such a state. Both variations imply the relevance of the mental element and do not seem to be of any particular value with respect to cyber crimes.

An interesting doctrinal debate dating back more than a century was aimed at clarifying the jurisdictional consequences of a libellous or an explosive letter sent from the territory of State A to the territory of State C, where it realises its offensive purposes, after having travelled through the territory of State B.[7] The concept of transit of digital data through the territory of a state (and its gateways, net, nodes, servers and even communication satellites as space objects) could be actualised in order to affirm that in such a state a material part of the conduct/ effect of the crime has taken place and in order to trigger its jurisdiction in a wider sense. The latter could be identified as a suggested trend in the development of the jurisdiction to adjudicate, with

5 Article 10 of the Cybercrime Convention refers to infringement of copyright, as defined under the law of that Party, pursuant to the obligations it has undertaken under the Paris Act of 24 July 1971 revising the Bern Convention for the Protection of Literary and Artistic Works, the Agreement on Trade-Related Aspects of Intellectual Property Rights and the WIPO Copyright Treaty and International Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organisations (Rome Convention).

6 References on the relevant documents in Lothar Bergmann, *Der Begehunsort im internationalen Strafrecht Deutschlands, Englands und der Vereinighten Staaten von Amerika*, Walter de Gruyter & Co., Berlin, 1966, p. 7ff.

7 The question on the relevance, for the determination of the *locus commissi delicti*, of that part of the conduct which, due to the intervention of other individuals (and also non human intervention), may further the conduct to its effect, is raised by Friedrich Meili, *Handbuch des internationalen Strafrechts & Starfprozessrechts*, Orell Füssli, Zürich, 1910, p. 119.

respect to crimes committed in or through cyberspace. It could also eventually be identified in the transit of data instrumental to a crime (eventually autonomous and not linked to the above referred broad definition of the *locus commissi delicti*) and also the prorogation of jurisdiction with respect to the violation or elusion of orders prohibiting or banning certain content from websites. Nevertheless, it should be taken into account that 'libel jurisdiction' in so called 'common law' systems currently requires, with respect to web-related cases, a 'substantial publication' which may well be considered an effect requirement of the conduct. Obviously the active and passive personality principles with respect to the exercise of jurisdiction to adjudicate may also provide guidance for crimes committed in or through cyberspace.

As a result of the evocative descriptions of cyberspace as a 'non space' and a legal limb – a description we currently do not agree with – cyberspace could be qualified as a 'non foreign' territory, similar to those ancient regimes labelled as 'lawless territories' (and the politically incorrect version referred to as 'non-civilised territories') where the jurisdiction to adjudicate was asserted without those limits, relying on the criminal character of the conduct in the place where it took place.

In order to exercise the jurisdiction to adjudicate, either of the criteria for the exercise of the jurisdiction to enforce should be fulfilled in order to prevent the further continuance of the crime, to identify the authors of the crime, to identify the victim and material witnesses, to gather and secure evidence and to prevent the escape of those having committed the crime. As an alternative to jurisdiction to enforce, suitable international agreements (or agreements within the EU domestic legislation implementing EU legislation on judicial cooperation) may assist.

Jurisdiction to enforce is eminently territorial and is exercised with respect to individuals, as well as goods, which can be found within the territory of the concerned state. With specific reference to conduct taking place in or through cyberspace, the territoriality principle implies the possibility to seize (physical) servers and data stored on servers located in the territory of the concerned state. Further (almost) territorial enforcement may refer to the so called expedited preservation of stored computer data,[8] the expedited preservation and disclosure of traffic data[9] and the 'production order' through a person or more frequently an ISP, even if the data are stored on a support physically located elsewhere as 'data in transit' until downloaded[10]. An enforcement activity with potentially extraterritorial reach is represented by the securing of information through an access point to the web.

The practice of telecommunications taping shows that, due to technical reasons, in specific circumstances the territorial state may not be in a condition to intercept a target within its territory and may need to rely on a third state which is able to enforce the measure. The latter state has a reduced interest in exercising supervisory jurisdiction, as the target is not in its territory and acceptance may be presumed after the expiry of a short notice. These issues have been partially addressed in the *Convention established by the Council in accordance with*

8 Such measures are established under art. 16 of the Cybercrime Convention.
9 Such measures are established under art. 17 of the Cybercrime Convention.
10 Search and seizure activity under article 19, paragraph 1, of the Cybercrime Convention, mentions the 'same territory' requirement only in respect of 'computer-data storage medium in which computer data may be stored' (lett. b) and not also in respect of a computer system or part of it and computer data stored therein (lett. a). The Explanatory Report § 192 states that 'the reference to "in its territory" is a reminder that this provision, as all the articles in this Section, concern only measures that are required to be taken at the national level'.

article 34 of the Treaty on European Union, on Mutual Assistance in Criminal Matters between Member States of the European Union, produced in Brussels on the 29th of May 2000,[11] article 18, 2, lett. *c*) .

Remote (live) access to information located on a physical support in the territory of a foreign state may to some extent infringe the territorial sovereignty of such a state. Remote (live) access represents one of the enforcement measures, to some extent similar to the traditional physical search of property.

In the recent past the German Constitutional Court confronted the issue of so-called 'online search' of hard discs, authorised under police laws of one of the German *Länder* for preventative purposes[12]. The Court set aside the law due to the inadequate definition of the prerequisites for the measure and deemed the constitutional provisions of measures impacting on communications rather than those on home search applicable. The judgment opened the way to several laws, including federal law.

Current practice shows that hard disc searches were performed through a trojan which infected a specific information system rather than through a remote (live) search. Furthermore, chronicles suggest that not only devices targeted with a specific authorised measure were infected, but in some case, devices were infected directly by customs when imported, by creating a *de facto* backdoor.[13]

This practice is questionable under the rule of law principle, as at least preparatory measures for a hard disc search and further measures, such as the parallel sending of calls to other unwanted recipients or the unwanted video-taping and photographing through the device, are adopted without judicial oversight even if the early infection of devices does not represent *per se* an interference with the individual who later purchases the device. Pre-installing a trojan in a device which will later connect with law enforcement activities instead of remote (live) access to a physical storage device in the territory of the state may determine a more vague interference with the jurisdiction and sovereignty of the foreign state where the device may later be located, as the interference may be considered a *de facto* effect of a previous law enforcement activity.

One aspect still to be examined is represented by the possibility that infection of information technology devices through trojans results in cross-border implications; that is, the element of provisions of criminal law in states where the device or system is located or from whose territory it is sending out unwanted information. Unclear legal procedures in the state enforcing its laws via trojans may result in the review of the authorising procedures and the denial of a claim for the legitimacy of the measure.

Apart from possible extraterritorial implications of the use of trojans as a law enforcement tool, state hacking methods, which are already a reality, imply the development of new patterns for

11 In EU Official Bulletin, C 197, 12th of July 2000: On the provisions dealing with telecommunications, *See* Barbara Huber, *Forschungsprojekt §12 FAG und Überwachung der Telekommunikation*, in, Wolter – Jürgen – Schenke, *Zeugnissverweigerungsrechte bei (verdeckten) Ermittlungsmassnahmen*, 2002, p. 61ff.

12 Marie-Theres Tinnefeid, *Online-Durchsuchung: Menschenrechte vs. virtuelle Trojaner*, in MMR, 2007, n. 3, p. 137ff.

13 Claim of the legal firm AFB *(Strafanzeige gegen den Einsatz des „Bayerntrojaners' gegen Staatsminister Joachim Herrmann, LKA-Präsident Peter Dathe sowie weitere Personen)* on the 17th of October 2011.

the judicial oversight of law enforcement activities which effectively take into account the risk of abuse by law enforcement agencies, as well as the risk of misuse by other subjects.

These patterns for judicial oversight should prevent systematic and mass infection of devices, establish an expiration date for pre-implanted backdoors, ensure proper and independent expertise on the part of the authorising judge, include inhibitory actions for those allegedly affected and, finally, establish proper liability mechanisms for those damaged by the measure.

In order to prevent conflict of jurisdiction in the form of conflicting decisions as to the legitimacy resorting to trojans for the surveillance of information technology devices, new international instruments should be developed and should refer to a mutual recognition of surveillance measures, based on information and access sharing. Further enforcement mechanisms could include tagging of IP numbers and real-time transfer to territorially competent authorities, as well as temporary blocking of the device used for violations until intervention by competent authorities.

Improvement of cross-border law enforcement may come from a reinforcement of the role of ISPs in enforcement activities, bearing in mind that, according to provisions already agreed within the EU Treaty on judicial assistance (art. 19), *'systems of telecommunications services operated via gateway on the territory, which for the lawful interception of the communications of a subject present in another State are not directly accessible on the territory of the later, may be made directly accessible for the lawful interception by that Member State through the intermediary of a designated service provider present on its territory'*.

Cross-border issues could perhaps be ameliorated by requiring multinational companies to preventatively agree (when authorised to operate) to execute requests for the storage, retrieval and seizure of data stored or accessible by them, even if the storage device is located in the territory of another state, and to develop 'standard service clauses' reserving them the right to execute foreign requests from the consumer.[14]

Beyond the above-mentioned contractual practice of cross-border cooperation, a much wider extent of multinational companies offering internet services should be included, in order to empower them to directly fulfil law enforcement tasks or at least delegated investigations with law enforcement purposes. The idea is to foster repression of cyber crimes through private actors acting as Private Law Enforcement Companies (PLEC) across state borders, throughout the company and its affiliates' reach, seeking (when needed under territorial criminal procedure law as *lex loci actus*) authorisation from judicial authorities and cooperating with prosecution offices for the repression of crimes, under territorial (for single act) or process (if prosecution starts) roles for oversight and liability. From an econometric perspective, requiring a contribution from those making money out of services in cyberspace for the repression of cyber criminality seems an acceptable onus and would justify budgets.

A topic partially related to the previous, and perhaps of more urgent character, is represented by

14 It should be noted that consent is currently a pre-requisite for the so called '*Trans-border access to stored computer data with consent or where publicly available*' under article 32 of the Cyberspace Convention.

'private' reactions to cyber attacks,[15] which may include the detection of intruders, disruption of attacks (by neutralising programmes as well as hardware) and gathering of information useful for the prosecution of those having committed the crime. Currently, intrusion detection may have cross-border implications and may trigger conflicting laws involving legal consequences. Far from advocating the right of companies targeted by cyber attacks to conduct 'private wars', there is the need to adopt uniform rules as to what represents a legitimate reaction under criminal defence and eventual liability patterns.

3. RIGHT TO ACCESS CYBERSPACE

The current social and democratic function of cyberspace is barely reflected in current human rights instruments and modern constitutions.[16]

In the cyberspace domain, the individual may express their personality, but the right to do so is properly defined in the negative, as the personality should not be affected by the fear of being subject to profiling through data collection and, in a wider sense, the individual should be granted the right to informational self-determination through a proper data protection regulation. The latter right is shown to be often affected in cyberspace by the acceptance of foreign data protection regulations offering a lower level of protection. There is, in this sense, a need for more homogenous regulations. Non-viable alternatives are represented by banning or restricting transactions, implying insufficiently strict data protection under foreign data protection rules.

Recent events show that the disclosure of personal information to authoritarian governments may lead to consequences under tort laws, while cooperation in order to implement censorship measures does not appear, currently, to be successfully challenged in court in order to obtain redress.[17]

A right for the individual to access cyberspace as a minimum social right is not currently recognised and taxes or fees are legitimate. Excessive taxes or fees could be considered a restriction to the right to access and provide information. From the perspective of free-trade agreements, taxes and fees, as well as limitations, may fulfil the requirements of a restriction to free trade and determine liability under WTO legal instruments.[18]

The so-called 'new media' which have developed on the World Wide Web, challenging traditional

15 For this purpose, See US National Research Council, *Technology, Policy, Law and Ethics Regarding US acquisition and use of cyberattack capabilities*, The National Accademies Press, Washington DC 2009, p. 77ff.

16 Even if focused on a state's responsibility in respect of cyber security, the *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*, A/66/359, deals in the annex *International code of conduct for information security*, with aspects related to the right to access cyberspace.

17 Michael Kwan, Kam-Pui Chow, Pierre Lai, Frank Law & Hayson Tse, *Analysis of the Digital Evidence Presented in the Yahoo!* Case, in IFIP Advances in Information and Communication Technology, Volume 306, 2009, p. 252ff.

18 Cynthia Liu, *Internet censorship as a trade barrier: a look at the WTO consistency of great firewall in the wake of the China – Google Dispute, in Georgetown Journal of International Law*, 2011, p. 1199ff.; Ritika Patni & Nihal Joseph, *WTO Ramification of Internet Censorship: The Google – China's Controversy in NUJS Law* Review, 3, 2010, p. 337ff.

media and specifically newspapers, derive the recognition of their roles under human rights and even democratic roles to include limiting restrictions from those media they are progressively replacing to those that are necessary in a democratic society. Specific preventative controls, counterbalancing in some countries the prohibition of any form of censorship and seizure of newspapers, may sometimes only work due to the parallel existence of the traditional media.

Social networks (such as *Twitter* and *Facebook*) which have proven to be drivers of social unrest in the so called 'Arab Spring' have also proven to represent a challenge for traditional media which, in an attempt not to be out-weighed by this form of new media, more often rely directly on such sources, while the traditional professional control over the reliability of information has become almost impossible. From this perspective, one could question if the individual's 'right to be informed' has, to a certain extent, been infringed by the speeding-up of information on one side and the information overflow on the other, while direct access to information by itself offers no guarantee of reliability, creating an illusion of access to first-hand news. The reverse side of the 'right to be informed', the 'right to inform' about relevant facts, has apparently overcome the tradition of freedom of speech for all and the role of traditional media, in favour of the faculty to post almost everything.

Apparently, social networks have recently played the role of drivers for democracy and have the potential to allow individuals to participate in not only political but also social and cultural life. However, they may also shelter discriminatory practices[19] including exclusion practices, the lack of protection against harmful content and the misuse of personal data.

The so called 'Arab Spring' shows that social media may to a certain extent fulfil the role of a command and control system in times of insurgency. The latter aspect should trigger the question of the real role of social networks as an efficient instrument for media and information operations and further, the role of leading multinational companies operating as non-state actors, which may be motivated by more than just profit-making purposes.

The risk of mass manipulation, which is *per se* one of the most difficult to counter with democratic means, has grown to a dimension which may no longer be managed or even mitigated by a single state. Perhaps governance mechanisms and social network management ethics could help in assuring that such instruments remain a driver for democracy rather than a means for non-conventional warfare.

Twitter has also shown an impact on legitimate law enforcement activities: the federal prosecution office of Brazil has requested an injunction to stop *Twitter* users from alerting drivers to police roadblocks, radar traps and drink-driving checkpoints. Such an injunction could make Brazil the first country to take *Twitter* up on its offer to censor content at governments' requests.[20]

Cyberspace has also become the domain in which information for educational purposes has become freely available and has become functional to the right to education of the individual, which has been universally recognised since the 1948 *Universal Declaration of Human Rights*

19 For this purpose we would like to recall the *Draft Recommendations of the Committee of Ministers to member states on the protection of human rights with regard to social networking services*, MC – NM (2010).003Final, of the Committee of Experts on New Media (MC – NM) of the Council of Europe contained in the document, adopted on 30th of November 2011 at Strasbourg.

20 Stan Lehman, Associated Press, 10 February 2012.

(UDHR, article 26). This has happened in an often spontaneous and non-institutionalised way, but often at the cost of copyright for material placed on the web.

The link between the internet and the right to education has been stressed by the Internet Rights & Principles Coalition,[21] which has attempted to define the implications of such rights on the internet. Organised attempts to make culture available through sites like *Google Books* have been ascertained as a violation of copyright and *Wikipedia* could also be affected by claims relating to the violation of copyright. In general terms, intellectual property rights are ensured in ways compatible with the aim to also grant the social function of such rights. Simply qualifying the posting of certain partial and limited contents covered by copyright on the web as a 'publication', despite proper quotation, may not properly balance copyright with the widespread educational purposes fulfilled by cyberspace. Accordingly one could question if, in this case, as in the case of advanced technology being beneficial to the wellbeing of mankind, a rule of non-excessive protection[22] of rights could apply in order to emphasise education and access to cultural aspects.

In cyberspace, access to information for any purpose is significantly influenced by search engines, responding to searches based on keywords. The result of the search is shown, as is well-known, as a hit-based list. Such an outcome may be influenced, and the hit list may neglect undesired content generally or based on the IP address of the individual making the search, supporting censorship mechanisms of authoritarian governments. The so called 'great firewall' developed for the Chinese market could also easily be used in democratic contexts. Such a threat also has an economic and anti-trust dimension, as *Google* could skew search results to favour its own services, making it hard for other businesses to win top advertising placements. *Google* came under the lens of data protection authorities not only for violations associated with *Google Street View*, *Google Earth* and *Google Maps*, but also in respect of the new social platform network *Google Buzz*. Besides, IP-associated storage of searches offers a unique potential for individual profiling and is correspondingly a unique threat to the individual's informational self-determination.

As in the case of social media, search engines have a unique potential for mass manipulation, and the multinational companies owning them have a dimension which allows them to outplay a single state. As such, the introduction of open oversight, governance mechanisms and company ethics should be promoted, oversight which should not be limited to a specific sector such as data protection or trade, but should cover a wide spectrum of all issues which may be associated with firewalls and search algorithms sensitive to democratic values. Mass manipulation may endanger democracy, but reference to the fear of mass manipulation evokes the risk of censorship and remedies which may be worse than the risk itself. Nevertheless, neglecting the risk no longer seems acceptable.

21 On the history of the development of the initiative, *See* Wolfgang Benedeck, Matthias C. Kettemann, Max Senges, *The Humanization of Internet Governance: A Roadmap towards a Comprehensive Global (Human) Rights Architecture for the internet*, Third Annual GigaNet Symposium, 2 December 2008, Hyderabad, India.

22 The debate on 'excessive protection' of intellectual property refers currently to protection of patents in agriculture, biotechnology, medicine and even ultra-high technology, which may have the effect of increasing economic and social inequality if the patents covering development are not made available as a consequence of excessive rights on such patents. Copyright could, in the authors' views, benefit from the debate as the Internet has intrinsically increased the need to access and disseminate the content of copyrighted works.

4. CONCLUSIONS

As the world became too small, some started dreaming and writing about virtual and infinite worlds that they could navigate without being affected any longer by daily problems. Suddenly they felt that the result of putting together internet service providers (ISPs), connecting computers to the internet and browsing websites maintained by web hosting services, was the emergence of a romantic new domain, global like no other, common to mankind and also border-free.

Perhaps the 'new romantic' view of cyberspace is misleading for the development of a clear vision from a legal perspective. Some of the problems posed by cyberspace are not really new, and resorting to ancient ideas may often appear to be beneficial. This could be true for those theories developed in order to clarify the role of transit of criminal instruments through the territory of a state.

Cooperation amongst authorities is often a matter of sovereignty and pride. Contractual development could foster judicial and police cooperation through the further development of contractual practice aimed at exploiting the potential role of multinational companies in the communications and IT sectors as an entry point for cross-border enforcement of requests. Further development could include the necessarily conventional development of a role for law enforcement functions to be carried out by private entities within multinational companies.

Obviously the reinforcement of the role of multinational companies presupposes the establishment of effective oversight mechanisms, also aimed at overcoming identified gaps.

Chapter 2
Cyber Policy & Strategic Options

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Russia's Public Stance on Cyberspace Issues

Keir Giles
Conflict Studies Research Centre
Oxford, UK
keir.giles@conflict-studies.org.uk

Abstract: Russian views on the nature, potential and use of cyberspace differ significantly from the Western consensus. In particular Russia has deep concerns on the principle of uncontrolled exchange of information in cyberspace, and over the presumption that national borders are of limited relevance there. Circulation of information which poses a perceived threat to society or the state, and sovereignty of the "national internet", are key security concerns in Russia.
This divergence undermines attempts to reach agreement on common principles or rules of behaviour for cyberspace with Russia, despite repeated Russian attempts to present norms of this kind to which other states are invited to subscribe.
This paper examines aspects of the two most recently released public statements of Russian policy on cyberspace: the "Draft Convention on International Information Security" (released 24 September 2011) and the Russian military cyber proto-doctrine "Conceptual Views on the Activity of the Russian Federation Armed Forces in Information Space" (released 22 December 2011) in order to describe the Russian public stance on cyberspace. Conclusions are drawn from the "Conceptual Views" on how the Russian Armed Forces see their role in cyberspace. The documents are referenced to the Information Security Doctrine of the Russian Federation (2000) as the underpinning policy document prescribing Russia's approach to information security overall, including its cyber elements.
The Russian authorities considered that protests over the State Duma election results in December 2011 arose at least in part because of a cyber/information warfare campaign against Russia. The informational and political response of the Russian authorities to this is taken as a case study to measure the practical impact of the Russian views outlined above. In addition, the dynamics of the London International Conference on Cyberspace are referenced in order to illustrate failure to achieve dialogue over the difference of these views from the Western consensus.

Keywords: *Russia, information security, social media, civil protest, policy, military*

1. INTRODUCTION

To external observers, dialogue between Russia and Western partners on cyberspace issues seems characterised by mutual incomprehension and apparent intransigence. Norms which are taken for granted on one side are seen as threatening by the other, and the lack of a common vocabulary or common concepts relating to cyberspace means that even when attempts are made to find common ground, these attempts soon founder.

According to Russia's Communications Minister Igor Shchegolev, "for the time being, in the West not everybody always understands what rules we are following" [1]. This remains true despite the fact that Russia has for over a decade been attempting to gather international support for these rules in a variety of international fora including the United Nations [2] and others [3].

This paper reviews two of the most recent public statements of the Russian approach to information security, a concept which carries cyber security implicitly within it, in order to extract key principles of the Russian approach. It then measures these principles against official and unofficial Russian state action against protest movements following the parliamentary elections in December 2011.

2. THE DRAFT CONVENTION

In September 2011, a "Draft Convention on International Information Security" was released at an "international meeting of high-ranking officials responsible for security matters" in Yekaterinburg, Russia, narrowly post-dating the "International Code of Conduct for Information Security" presented by Russia and other states at the United Nations [4].

The key provisions of the document have been condensed into a list of 23 fundamental issues of concern to Russia in information space by the Institute of Information Security Issues (IISI) of Moscow State University, which is closely engaged in developing the draft Convention. These issues, each of which is reflected in one or more articles of the proposed document, include some provisions which should excite no controversy in any part of the world, such as avoidance of breaches of rights and freedoms, or "criminalisation of use of information resources for illegal purposes". But at the same time, a number of the issues raised run counter to the views on use and governance of the internet that have emerged in the USA, UK and other like-minded states – a system of views which forms an unstated but nonetheless tangible concurrence - referred to further, for brevity and clarity, as "the Western consensus". This consensus, while regularly voiced at international events like the London International Conference on Cyberspace on 1-2 November 2011, is also expressed in a number of published international documents, for example the Organisation for Economic Cooperation and Development (OECD) recommendations on principles for internet policy making released shortly afterwards [5].

A key divergence between Russian and Western approaches to cyber security is the Russian perception of content as threat [6]. In the Russian list of issues of concern, this is expressed as the "threat of the use of content for influence on the social-humanitarian sphere". By contrast,

the Western consensus recognises the threat from hostile code, but generally discounts the issue of hostile content. The OECD recommendations referred to above, for example, include

> free flow of information and knowledge, the freedom of expression, association and assembly, the protection of individual liberties, as critical components of a democratic society and cultural diversity [5]

It is regularly stated as a fundamental principle "that cyberspace remains open to innovation and the free flow of ideas, information and expression", as stated by UK Foreign Secretary William Hague and others at the London Conference referred to above [7]. Yet at the same conference, Minister Shchegolev attached important caveats to the principle of free flow of information: this should be subject both to national legislation, and to counter-terrorism considerations - chiming with another principle on the list, "restrictions of rights and freedoms only in the interests of security" [8].

Thus while both sides publicly espouse the freedom of exchange of information, and thus occasionally give the illusion of consensus, the Russian reservations on how far this principle can safely be extended mean that in practical terms the two views are as far apart as ever.

Two further issues identified by IISI, "Refraining from using information and communications technology to interfere in the affairs of other states" and "Threat of use of a dominant position in cyberspace" lie behind the perception voiced by certain sections of the Russian leadership that protests following the parliamentary elections in December 2011 were inspired, facilitated and financed from abroad - to be discussed further below. In particular, the mention of a "dominant position in cyberspace" refers to the idea of "information space [being] a place of competition over information resources... The USA is currently the only country possessing information superiority and the ability significantly to manipulate this space [9]."

The principle of indivisibility of security is highlighted in the draft Convention. Here again, apparent consensus hides fundamental disagreement - simply because this common phrase has entirely different meanings in Russian and in English. Despite recognition and patient explanation that use of the identical phrase to refer to widely differing concepts leads to misunderstanding and frustration [10], the phrase continues to occur in both Western and Russian discourse leading to each side embarking on their own separate conversation [11].

"Internet sovereignty" is another key area of disagreement. Russia, along with a number of like-minded nations (for example members of the CIS, CSTO and SCO), strongly supports the idea of national control of all internet resources that lie within a state's physical borders, and the associated concepts of application of local legislation - or as worded in the draft Convention itself, "each member state is entitled to set forth sovereign norms and manage its information space according to its national laws" (Article 5.5). This is in direct opposition to the approach of, for example, the USA, as expressed firmly by US Secretary of State Hillary Clinton in December 2011, saying that countries like Russia wished to

> empower each individual government to make their own rules for the internet that not only undermine human rights and the free flow of information but also the interoperability of the network. In effect, the governments pushing this agenda want to create national barriers in cyberspace. This approach would be disastrous for internet freedom [12].

The list of underlying principles provided by IISI includes "Taking essential measures to prevent destructive information activity from territory under the jurisdiction of a state". This vaguely-worded but ominous-sounding provision refers to a section in the draft Convention which covers states ensuring that information infrastructure within their own jurisdiction is not used for hostile activity, and cooperating in order to identify the source of such activity. (Article 6.2). Consideration of the practical implications of a stipulation of this kind, and the obligations it entails, leads quickly to the realisation of an enormous legislative and administrative burden on states which might wish to subscribe to the draft Convention. Not only must they supervise the legality of content within their own jurisdiction, but also ensure that it is considered inoffensive and non-hostile in the jurisdictions of all other signatories – otherwise, they can immediately be accused of permitting hostile activity in breach of the Convention.

Another key stipulation which is gravid with misunderstanding is the provision for "taking measures of a legal or other nature which are essential for access with grounds and in a legal manner to specific parts of the information and communications infrastructure of a State Party". In the current text of the draft Convention, this appears as "take necessary steps of legislative or other nature which will guarantee lawful access to specific parts of the information and communication infrastructure in the territory of the State Party which are legally implicated in being employed for the the perpetration of terrorist activities in information space" (Article 9.5).

Two important areas of conceptual divergence arise here: first, the mention of "terrorism", and second, the issue of access to a foreign state's information space.

Conceptual differences in the understanding of the nature of "terrorism" between Russian and other states provide an additional layer of complexity and indeterminacy to the already muddied picture of what constitutes "cyberterrorism". As described by Anna-Maria Talihärm [13], Alex Michael [14] and others, "there is a great abundance of different definitions of the idea of 'terrorism'... the addition of the prefix "cyber" has only extended the list of possible definitions and explanations".

Thus without consensus with Russia on what precisely is covered by "perpetration of terrorist activities in information space", this clause remains unusable. Such consensus is unlikely to be achieved given the fundamental and unresolved differences between the two sides on what constitutes both terrorism and counter-terrorist activity [15].

At the same time the call for authorised access to information infrastructure in another state's jurisdiction is reminiscent of the text of Article 32 of the Council of Europe Convention on Cybercrime (the Budapest convention):

> A Party may, without the authorisation of another Party... access or receive, through a computer system in its territory, stored computer data located in another Party, if the Party obtains the lawful and voluntary consent of the person who has the lawful authority to disclose the data to the Party through that computer system [16].

- yet this text constitutes Russia's main objection to ratification of the Budapest convention [17]. The key phrase which prompts Russian objections is "without the authorisation of another Party". In the Russian view, this is an intolerable infringement on the principle of sovereignty as described above. In addition, the range of options covered by "the person who has the lawful authority to disclose the data" is a source of concern, including as it may organisations other than the State. Russian concerns over practical application of the Budapest convention are illustrated by a report in the official government newspaper which highlighted the "dubious provision for foreign special services to invade our cyberspace and carry out their special operations without notifying our intelligence services" [18].

In sum, then, the articles of the draft Convention and its underlying principles serve well to illustrate the two emerging consensuses on governance of the Internet: the Western one, insisting on the free, unrestricted and ungoverned flow of information, and the consensus espoused by Russia and like-minded states, with important caveats on the flow of information and an insistence on national sovereignty in cyberspace.

3. "CONCEPTUAL VIEWS"

The most recent official Russian policy statement on cyber issues to be published at the time of writing is the "Conceptual Views on the Activity of the Russian Federation Armed Forces in Information Space". This document was presented at an information security conference in Berlin on 14 December 2011 [19], and released in text form on 22 December 2011 [20].

Despite a large volume of previous semi-official literature on information warfare, this is the first explicit public statement of the Russian military's role in cyberspace, and has been described as a Russian military cyber proto-doctrine. When compared to similar documents released in the USA, UK and elsewhere, it is as interesting both for what it includes and for what it omits.

This is a specifically Russian document, and does not resemble its foreign counterparts, for example the US Department of Defense Strategy for Operating in Cyberspace [21] - not only through references to supporting doctrinal documents (the Military Doctrine and Information Security Doctrine of the Russian Federation) but also in its underlying presumptions and definitions of information challenges.

In this way it reflects a long-standing recognition not only that potential operations in information space pose an entirely new set of challenges [22], but also that foreign concepts of information security, along with those of other areas of military endeavour, are not applicable to Russian circumstances - as expressed in 1995 by prominent Russian military commentator Vitaliy Tsymbal:

> It is false to presume that we can expediently interpret and accept for our own use foreign ideas about information warfare (IW) and their terminology in order to avoid confusion and misunderstanding at international discussions, during information exchanges, or during contact between specialists. Quite the opposite, it makes no sense to copy just any IW concept. Into the IW concept for the Ministry of Defence of the Russian Federation (RF) must be incorporated the constitutional requirements of the RF, its basic laws, specifics of the present economic situation of the RF, and the missions of our Armed Forces [23].

With the exception of references to the economic situation, this is precisely what the Views have done.

They echo the defensive theme of other Russian documents relating to cyberspace, including the draft Convention described above, and cite in their preamble a statement of the external threat to Russia's information security arising from other states developing information warfare concepts [6]. Further, they state that "a targeted system of activity has been established in the Armed Forces of the Russian Federation intended to provide for effective deterrence, prevention and resolution of military conflicts in information space".

The definition of the information war which the Armed Forces are called upon to deter and prevent is worth citing in full, as it illustrates the enduring holistic nature of the Russian perception of information warfare and cyber conflict as an integral part of it. Information war, according to the Views, is

> "conflict between two or more states in information space with the aim of causing damage to information systems, processes and resources, critically important and other structures, subverting the political, economic and social systems, **mass psychological work on the population to destabilise society and the state**, and coercing the government to take decisions in the interests of the opposing side." (Section 1, Fundamental Terms and Definitions - emphasis added.)

Legality (or, we should say, conforming with Russian law and international law as interpreted by Russia) is emphasised as the first principle governing military activity. Along with customary references to the primacy of international law, and the principle of non-interference in the internal affairs of other states, the Views note that use of the Armed Forces outside the Russian Federation is subject to a process of Federal Assembly approval, and states that "this provision should also be extended to the use of the Armed Forces of the Russian Federation in information space". (Section 2.1, Legality.) The Views also make provision for "deploying forces and resources to provide for information security on the territories of other states" (Section 3.2, Resolving Conflicts.) – which leads progressively-minded non-military Russian internet experts to speculate wryly on the picture of "commandos parachuting into server centres, iPads in hand".

The first priority for the Armed Forces is stated as "striving to collect current and reliable information on threats" and developing countermeasures - but this is explicitly for military purposes. The aim is primarily to protect military command and control systems and "support

the necessary moral and psychological condition of personnel". This has become essential since "now hundreds of millions of people (whole countries and continents) are involved in the unified global information space formed by the internet, electronic media and mobile communications systems". What is absent is mention of a military role in assessing or countering threats to broader society or the Russian state. (Section 2.2., Priorities.)

Russian military activity in information space "includes measures by headquarters and actions by troops in intelligence collection, operational deception, radioelectronic warfare, communications, concealed and automated command and control, the information work of headquarters, and the defence of information systems from radioelectronic, computer and other influences". In common with other Russian public statements, and in contrast to similar statements from other nations [24] and overt preparations by those states [25], what is absent from the Views is any mention of offensive cyber activity. (Section 2.3, Complex Approach.)

Also in contrast to foreign doctrinal statements, the Views list "the establishment of an international legal regime" regulating military activity in information space as the main aim of international cooperation with "friendly states and international organisations". (Section 2.5, Cooperation.)

These friendly organisations are later defined: the priorities are the Collective Security Treaty Organisation (CSTO), the Commonwealth of Independent States (CIS) and the Shanghai Cooperation Organisation (SCO). But these are groups of states which have already made substantial progress in formalising their shared views on information security; views in line with those of Russia as described earlier in this paper. The CSTO has a "Program of joint actions to create a system of information security of the CSTO Member States" [26] while the SCO has concluded an "Agreement among the Governments of the SCO Member States on Cooperation in the Field of Ensuring International Information Security" [27,6].

But in addition to this, the military are supposed to "work for the creation under the United Nations of a treaty on international information security extending the remit of commonly-accepted norms and principles of international law to information space". The Russian military is thus intended to have an explicit political role in promoting initiatives like the draft Convention on International Security referred to above, beyond simply having a voice in their drafting or having places on delegations; not a role which would sit naturally with most Western militaries.

This emphasis on international legal efforts echoes statements made by senior Russian military figures following the armed conflict with Georgia in August 2008. General Aleksandr Burutin, at the time Deputy Chief of the General Staff, said that the General Staff had recommended the development of an international mechanism to hold states to account for beginning information warfare, and furthermore that it was necessary "to move from the analysis of challenges and threats in information security to response and prevention" [28].

Both of these aspirations are reflected in the Views, and the intention to hold states to account for activity perceived as hostile which emanates from their territory is also reflected in the draft Convention as described above.

4. THE INFORMATION SECURITY DOCTRINE

Both of the documents described above make reference, either explicitly or implicitly, to the Information Security Doctrine of the Russian Federation (2000) [29].

This "doctrine", in the Russian sense of "national policy", is the fundamental document governing Russia's approach to information security, and as an integral subset of information security, cyber issues. It appears at first sight to contain the same liberal provisions for free exchange of information as called for by William Hague and Hillary Clinton as cited above. It is intended, inter alia, to "ensure the constitutional rights and freedoms of man and citizen to freely seek, receive, transmit, produce and disseminate information by any lawful means". (Article I, Part 1) It is only on closer inspection that the divergences with Western concepts and practices become clear.

A prime example lies in treatment of the media, whether state-owned or independent. The Doctrine stipulates "development of methods for increasing the efficiency of state involvement in the formation of public information policy of broadcasting organizations, other public media" (Article I, Part 4). The underlying concept, reflected in other doctrinal statements, is that media are a tool of the state for shaping public opinion in a manner favourable to the authorities. As tellingly explained by one leading Russian security specialist in the Ministry of Defence's "Red Star" newspaper:

> How can you successfully wage an information struggle if during [conflict in] Chechnya a significant part of the mass media is taking the side of the specialists? We need a law on information security [30].

- the implicit assumption being that information security must necessarily involve ensuring that the views transmitted by media, independent or not, are favourable to the government.

At the time of the release of the Information Security Doctrine, Col-Gen Vladislav Sherstyuk, then First Deputy Secretary of the Security Council of the Russian Federation responsible for information security and one of the key drafters of the document, explained that the doctrine would not be used to restrict independent media, but that nonetheless all media, government or private, must be under state supervision [31]. At the same time the visceral reaction of some sections of the Russian leadership to dissenting views voiced through independent media was evinced by the response of Prime Minister Putin to reporting on European missile defence plans by the Ekho Moskvy radio station: Putin described the experience of listening as "having diarrhoea poured over him day and night" [32]. How much more emphatic still must be the reaction of Putin, and those who think like him, to vitriolic online attacks on the current leadership via foreign-owned social media.

The Doctrine deals with issues such as these by stating that "the main activities in the field of information security of the Russian Federation in the sphere of domestic policy are … intensification of counter-propaganda activities aimed at preventing the negative effects of the spread of misinformation about the internal politics of Russia" (Article II, Part 6) as well as

"development of specific legal and institutional mechanisms to prevent illegal information-psychological influences on the mass consciousness of society" (Article II Part 7). Capacity for "preventing negative effects" was tested by online organisation of mass protest rallies following the elections to the Russian parliament on 4 December 2011.

5. CASE STUDY: INFORMATION WARFARE AGAINST RUSSIA?

The official and unofficial Russian responses to protest and dissent following the parliamentary elections appeared confused and contradictory. Interference with information resources was evident, but stopped short of the complete information blockade expected by some commentators [33].

The examples given above of doctrinal concern over the circulation of information should illustrate that the permissibility or otherwise of expressing or organising dissent in cyberspace is not clear-cut. Civil protests over the election results perhaps fell in a grey area for some security practitioners in Russia between legitimate protest and dangerous subversion, leading to a mixed response including brief and sometimes ineffectual attempts to block opposition communications and internet resources.

Suspicion of foreign involvement triggered fear of subversion and "colour revolution", linked to the pervasive Russian argument that political instability in North Africa and the Middle East resulted from the plotting of the West led by the USA [34]. In addition to the battery of colourful accusations on this topic from Russia's more hawkish senior commentators, President Medvedev echoed the view that Russia was vulnerable to the same kind of interference. Speaking in February 2011, he said:

> Look at the situation that has unfolded in the Middle East and the Arab world. It is extremely bad. There are major difficulties ahead... We need to look the truth in the eyes. This is the kind of scenario that they were preparing for us, and now they will be trying even harder to bring it about [35].

And indeed the progress of the NATO campaign in Libya only deepened the sense of alarm felt in Russia [36] - not least because the Libya campaign precisely matched the pattern for "modern warfare" described by Chief of General Staff Nikolay Makarov in published articles including one the previous year: "use of political, economic and information pressure and subversive actions, followed by the unleashing of armed conflicts or local wars, actions that result in relatively little bloodshed" in order to achieve the aggressor's intent [37].

Observing processes of this kind gives rise to two key concerns in Russia: first, the precedent set for interference in the internal affairs of a sovereign state with the intention of regime change; and second, the risk that intervention "could unpredictably lead to a large-scale war involving unforeseen adversaries" [37].

At the time of writing, both of these concerns are informing Russian objections to Western pressure on the Syrian government, most recently expressed in a Russian and Chinese veto of a UN Security Council resolution on 6 February 2012. But at least part of the threat perception appears to derive from mirror-imaging: projecting Russian views onto foreign partners, and assuming they proceed from motivations which appear logical and rational through a Russian prism.

As Tim Thomas points out in discussion of Russian information warfare techniques:

> Disinformation is a Russian technique that manipulates perceptions and information and misinforms people or groups of people. Some disinformation techniques are quite obvious, some are unconvincing, and others work through delayed perception, rumours, repetition or arguments. Specific persons or particular social groups can serve as disinformation targets... In Russia today, where an unstable public-political and socio-economic situation exists, the entire population could serve as the target of influence for an enemy disinformation campaign. This is a major Russian fear [38].

This fear gives rise to yet further incompatibilities between the Russian approach to internet freedom and that of other countries. At a U.N. disarmament conference in 2008 [39], a Russian Ministry of Defence representative suggested that any time a government promoted ideas on the internet with the intention of subverting another country's government, including in the name of democratic reform, this would be qualified as "aggression" and an interference in internal affairs [3]. This is immediately relevant to Russian suggestions that the USA was fostering and financing the post-election protests.

There appeared to be a coordinated campaign in response to the election protests, one neither avowed nor condemned by official Russian spokesmen. Distributed denial of service (DDoS) attacks were noted against election monitoring organisations and independent media, including against secondary targets that were reposting or hosting information from the primary list. With Twitter emerging as a key tool for organising rallies during December 2011 [40], Twitter activity by protesters was targeted for flooding by pre-positioned Twitter bots [41]. There was a formal request by the Federal Security Service (FSB) to the VKontakte social networking site to block specific pages organising protests, which was politely declined as illegal by VKontakte [42].

Yet this activity targeting opposition communications was brief in duration, and extended only a few days after the elections themselves; since when any repeat effort (at the time of writing, the most recent opposition protest of any significant size was on 4 February 2012) has been sporadic and on a much smaller scale.

One interpretation is that the Russian authorities wished to suppress communications but found the tools at their disposal to be limited. As described by analyst Kimberly Zenz, posting on LinkedIn in January 2012, "Targeting domestic sites didn't work, attempting to manipulate content on foreign sites didn't work, and domestic companies (LiveJournal and then VKontakte) did not prove to be reliable partners. Truly viable options for state management of online content appear to be lacking." This ties in with the commonly-held view that "the

swift emergence of the protests caught the government by surprise and revealed its inability to understand both the degree of discontent among the Russian urban population and the growing power of social media [43]."

The sense that the online protests were permitted, although not officially in favour, left state media falling back on interviews and features describing the evils of social media, including privacy concerns over Facebook [44] and incidents of suicide following cyber bullying [45], not to mention running articles by leading information warfare theorist Igor Panarin describing the foreign-backed information campaign against Russia [46].

Meanwhile the aspiration for control of the media described above resulted, among other things, in the issuing of clear instructions to the independent media on the right way to cover pro-Putin demonstrations - the "right way" including emphasising that those present are participating spontaneously and voluntarily, and not showing officials or official buildings [47].

Other elements of "intensification of counter-propaganda activities" as per the Information Security Doctrine included a retreat to more old-fashioned methods of tackling the opposition. A succession of dirty tricks was carried out at varying levels of competence and effectiveness, from frankly poor attempts at photo editing to discredit opposition figurehead Aleksey Navalny [48], through the publication of hacked e-mails from the Golos election monitoring organisation demonstrating that it received foreign funding (which Golos had not previously concealed) [49], to the release of telephone intercepts of veteran opposition leader Boris Nemtsov obscenely excoriating fellow opposition figures [50] and the planting of fake interviews with opposition figures in US media [51]. In March 2012, a documentary by NTV, a broadcaster with a long history of turbulent and shifting relations with officialdom and the official line, attracted widespread scorn online for its hostile portrayal of the protests, their participants and organisers [52].

The mixed response to online protests appears to reflect mixed views among the Russian leadership regarding the desired extent of internet regulation. In an article entitled "USA Hides Behind Fairy Tales About Human Rights", Secretary of the Security Council of the Russian Federation Nikolay Patrushev observed that some degree of internet regulation is essential. "Of course there should be reasonable regulation in Russia, just as it is done in the United States, China and many other countries," Patrushev wrote [53]. This chimed with the recommendation from Maj-Gen Aleksey Moshkov of the Interior Ministry's Bureau of Special Technical Measures (which includes Directorate K, responsible for dealing with cyber crime) that online anonymity should be restricted [54]. Meanwhile, among a range of other more ambiguous comments, Communications Minister Shchegolev stated uncompromisingly that "although cyber security and behaviour online are current problems in today's world, blocking the internet or restricting access to social networks is unacceptable under any circumstances". "There is an opinion that the Russian government is allegedly striving to achieve greater state control over the internet. But in Russia we are not even considering the possibility of blocking access to Twitter or Facebook, while in some European countries it has been openly stated that this will be done," he continued [1].

6. CONCLUSION

While informed by a substantially different world view from what is commonly accepted in the West, the Russian response to online dissent following the December elections was neither as draconian as sometimes portrayed in Western commentary, nor as liberal as a superficial reading of Russian policy documents would suggest. Russia will continue to push for international agreements regulating cyberspace, along the lines of the consensus already achieved with like-minded states in the CSTO and SCO. The challenge for any Western interlocutor seeking to engage with Russia on these issues is to understand that in cyber, as in so much else, the fundamental assumptions governing the Russian approach are very different from our own – and in many cases, similar language with divergent meaning employed by the two sides serves only to mask these differences.

REFERENCES:

[1] Interfax, "Shchegolev: tsenzury Interneta v Rossii ne dopustyat," 20 January 2011. [Online]. Available: http://www.interfax.ru/print.asp?sec=1448&id=226823.

[2] T. Maurer, "Cyber Norm Emergence at the United Nations," September 2011. [Online]. Available: http://www.un.org/en/ecosoc/cybersecurity/maurer-cyber-norm-dp-2011-11.pdf.

[3] T. Gjelten, "Seeing The Internet As An 'Information Weapon'," 23 September 2010. [Online]. Available: http://www.npr.org/templates/story/story.php?storyId=130052701.

[4] *International code of conduct for information security*, Annex to the 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 (A/66/359), 2011.

[5] OECD, "OECD Council Recommendation on Principles for Internet Policy Making," 13 December 2011. [Online]. Available: http://www.oecd.org/dataoecd/11/58/49258588.pdf.

[6] K. Giles, "Information Troops: A Russian Cyber Command?," in *Third International Conference on Cyber Conflict*, CCDCOE, 2011.

[7] W. Hague, "Chair's statement," 2 November 2011. [Online]. Available: http://www.fco.gov.uk/en/news/latest-news/?view=PressS&id=685663282.

[8] I. Shchegolev, in *London Conference on Cyberspace*, 2011.

[9] S. Modestov, "Prostranstvo budushchey voyny (The Space of Future War),," *Vestnik Akademii Voyennykh Nauk (Bulletin of the Academy of Military Science)*, No. 2, 2003.

[10] NDC, "The Indivisibility of Security: Russia and Euro-Atlantic Security," NATO Defense College, Rome, 2010.

[11] A. Monaghan, "NATO and Russia: resuscitating the partnership," May 2011. [Online]. Available: http://www.nato.int/docu/review/2011/NATO_Russia/EN/index.htm.

[12] H. Clinton, "Remarks by Hillary Rodham Clinton at Conference on Internet Freedom, The Hague, Netherlands," 8 December 2011. [Online]. Available: http://www.state.gov/secretary/rm/2011/12/178511.htm.

[13] A.-M. Talihärm, "Cyberterrorism: in Theory or in Practice?," *Defence Against Terrorism Review, Vol. 3, No. 2*, pp. 59-74, 2010.

[14] A. Michael, "Cyber Probing: The Politicisation of Virtual Attack," Defence Academy of the United Kingdom, Shrivenham, 2010.

[15] A. Monaghan, "The Moscow metro bombings and terrorism in Russia," June 2010. [Online]. Available: http://www.ndc.nato.int/research/series.php?icode=1.

[16] Council of Europe, "Convention on Cybercrime," 23 November 2001. [Online]. Available: http://conventions.coe.int/Treaty/EN/Treaties/Html/185.htm.

[17] V. P. Sherstyuk, *Presentation*, Brussels, 2011.

[18] T. Borisov, "Virtual'nyy mir zakryt," *Rossiyskaya Gazeta*, 12 11 2010.

[19] *Challenges in Cybersecurity - Risks, Strategies, and Confidence-Building*, Berlin, 2011.

[20] Russian Ministry of Defence, 22 December 2011. [Online]. Available: http://ens.mil.ru/science/publications/more.htm?id=10845074@cmsArticle.

[21] US Department of Defense, " Strategy for Operating in Cyberspace," July 2011. [Online]. Available: http://www.defense.gov/news/d20110714cyber.pdf.

[22] V. M. Lisovoy, "O zakonakh razvitiya vooruzhennoy bor'by i nekotorykh tendentsiyakh v oblasti oborony," *Voyennaya Mysl'*, no. 5, 1993.

[23] V. Tsymbal, *Concept of Information Warfare*, Moscow, 1995.
[24] D. Miles, "Doctrine to Establish Rules of Engagement Against Cyber Attacks," 20 October 2011. [Online]. Available: /www.defense.gov/news/newsarticle.aspx?id=65739.
[25] T. Miles, "Army activates first-of-its-kind Cyber Brigade," 9 December 2011. [Online]. Available: http://www.army.mil/article/70611/Army_activates_first_of_its_kind_Cyber_Brigade/.
[26] Collective Security Treaty Organisation, "CSTO website," 2012. [Online]. Available: http://www.odkb.gov.ru/start/index_aengl.htm.
[27] Shanghai Cooperation Organisation, 2009. [Online]. Available: http://www.sectsco.org/EN/show.asp?id=182.
[28] ITAR-TASS, 29 January 2009.
[29] Security Council of the Russian Federation, "Information Security Doctrine of the Russian Federation (2000)," 2000. [Online]. Available: http://www.scrf.gov.ru/documents/6/5.html.
[30] G. Miranovich, "Voyennaya reforma: problemy i suzhdeniya (Military Reform: Issues and Judgements)," *Krasnaya Zvezda*, 31 July 1999.
[31] Interfax, 12 October 2000.
[32] G. Novostey, ""I don't get upset with you when you pour diarrhoea on me": Putin chats with media leaders," 19 January 2012. [Online]. Available: http://www.city-n.ru/view/296196.html.
[33] Deutsche Welle, "Russia holding back online shutdowns for now, expert says," 13 December 2011. [Online]. Available: http://www.dw.de/dw/article/0,,15599135,00.html.
[34] A. Monaghan, "Flattering to deceive? Change (and continuity) in post election Russia," March 2012. [Online]. Available: http://www.ndc.nato.int/research/series.php?icode=3.
[35] D. Medvedev, "Dmitriy Medvedev provel vo Vladikavkaze zasedaniye Natsionalnogo antiterroristicheskogo komiteta," 22 February 2011. [Online]. Available: http://www.kremlin.ru/transcripts/10408.
[36] K. Giles, The State of the NATO-Russia Reset, Oxford: Conflict Studies Research Centre, 2011.
[37] N. Makarov, "Kharakter vooruzhennoy borby budushchego (The Character of Future Armed Conflict)," *Vestnik Akademii Voyennykh Nauk (Bulletin of the Academy of Military Science)*, 2010.
[38] T. Thomas, Recasting the Red Star, Fort Leavenworth: Foreign Military Studies Office, 2011.
[39] UNIDIR, 2008. [Online]. Available: http://www.unidir.org/audio/2008/Information_Security/en.htm.
[40] R. Soloveitchik, "Twitter Becomes Key for Moscow Protests," 23 December 2011. [Online]. Available: http://www.themoscowtimes.com/arts_n_ideas/article/twitter-becomes-key-for-moscow-protest s/450350.html.
[41] B. Krebs, "Twitter Bots Drown Out Anti-Kremlin Tweets," 8 December 2011. [Online]. Available: http://krebsonsecurity.com/2011/12/twitter-bots-drown-out-anti-kremlin-tweets/.
[42] Forbes Russia, "Durov: FSB prosit "VKontakte" blokirovat oppozitsionnye gruppy," 8 December 2011. [Online]. Available: http://www.forbes.ru/news/77291-durov-fsb-prosit-vkontakte-blokirovat-oppozitsionnye-gruppy.
[43] FIIA, *Finnish Institute of International Affairs seminar, "Russian Society through the Prism of Current Political Protests"*, Helsinki, 2012.
[44] Russia Today, "Stallman: Facebook IS Mass Surveillance," 2 December 2011. [Online]. Available: http://rt.com/news/richard-stallman-free-software-875/.
[45] Russia Today, "Social networks – a threat for Russia?," 2 January 2012. [Online]. Available: http://rt.com/news/social-networks-bullying-russia-695/.
[46] I. Panarin, "December 2011: Information War against Russia," 30 December 2011. [Online]. Available: http://rt.com/politics/information-war-russia-panarin-009/.
[47] Gazeta.ru, "Ne pokazyvat i ne upominat (Don't Show and Don't Refer)," 30 December 2011. [Online]. Available: http://www.gazeta.ru/politics/elections2011/2012/01/30_a_3979953.shtml.
[48] A. Kramer, "Smear in Russia Backfires, and Online Tributes Roll In," 8 January 2012. [Online]. Available: http://www.nytimes.com/2012/01/09/world/europe/smear-attempt-against-protest-leader-backfires-in-russia.html?_r=1.
[49] Zeenews, "Russian website publishes vote monitor's e-mails," 9 December 2011. [Online]. Available: http://zeenews.india.com/news/world/russian-website-publishes-vote-monitor-s-e-mails_746183.html.
[50] G. Faulconbridge, "Phone hacking Russian style: Opposition under fire," 20 December 2011. [Online]. Available: http://in.reuters.com/article/2011/12/20/russia-phonehacking-idINDEE7BJ0AE20111220.
[51] T. Lipien, "VOA harms Putin opposition in Russia," 8 February 2012. [Online]. Available: http://www.washingtontimes.com/news/2012/feb/8/voa-harms-putin-opposition-in-russia/.
[52] *Anatomiya Protesta*. [Film]. NTV, 2012.
[53] Argumenty i Fakty, "Nikolay Patrushev: SShA prikryvayutsya skazkami o pravakh cheloveka," 14 December 2011. [Online].
[54] M. Falaleyev, "Politseyskoye upravleniye "K" predlozhilo zapretit anonimnyye vystupleniya v Internete," 8 December 2011. [Online]. Available: http://www.rg.ru/2011/12/08/moshkov.html.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

French Cyberdefence Policy

Patrice Tromparent
Delegation for Strategic Affairs
Ministry of Defense
Paris, France
patrice.tromparent@intradef.gouv.fr

Abstract: Since 2008, France has initiated a proactive cyberdefence policy in order to remain one of the first nations in the cyber realm and to ensure its security. This policy testifies to the need for a global approach to cyber, which could be useful for countries trying to develop relevant frameworks and synergies to address the new challenges of cyberspace.
This article aims to describe and analyse this French official policy. It is based on up-to-date documents, most of them only available in French, and some not even published yet.
Every aspect of French cyber policy is taken into account, in particular the very specific mechanism to ensure the security of critical infrastructures. Indeed, France, which is an old centralised state, has built up a national cyberdefence authority which regulates not only the public sector, but also the private sector. Some other changes are also interesting to analyse: the ongoing process of transformation of the Ministry of Defence, and the complex links between public and private sectors. France also acts on the international stage, in particular within NATO and the EU, to build up multiple levels of cooperation between nations and to ensure a better regulation of cyberspace. In so doing, France has to reassess its traditional balance between national sovereignty and interdependence.
As a result, like many countries, France has to develop new concepts in order to address the global cyberspace challenges ahead as far as forms of sovereignty, legal and ethic issues and military operations are concerned, potentially bringing new opportunities for international cooperation.

Keywords: *France, cyberdefence, cyberstrategy, cyberpolicy*

1. INTRODUCTION

The first duties of a state are the protection of its citizens, the resilience of its society and economic and social progress. Communication and information systems have become the nervous systems of our modern society and are now essential for economic and social life. The French White Paper on Defence and National Security of 2008 states publicly that the security and defence of cyberspace are a priority:

> "France must retain its areas of sovereignty, concentrated on the capability required for the maintenance of the strategic and political autonomy of the nation: nuclear deterrence; ballistic missiles; SSBNs and SSNs; and cyber-security are amongst the priorities." [1].

Since then, France has initiated a proactive cyberdefence policy in order both to remain one of the first nations of the cyber world and to ensure its own security. The main evolutions are the ongoing implementation of a defence and security continuum, as well as the gathering of all the actors in order to address the multiform threats in cyberspace.

France also acts on the international scene to build up multiple levels of cooperation between nations and to ensure a better regulation of cyberspace. Cyberspace defence also raises questions about the new forms of sovereignty, the legal and ethical framework and military operations.

2. THE WHITE PAPER ON DEFENCE AND NATIONAL SECURITY AND CYBERSTRATEGY

Like many other nations, France publishes a global assessment of the geostrategic situation on a regular basis in order to determine the directions of major defence policy-making[1]. The White Paper on Defence and National security of 2008 identified for the first time cyberspace as a vital challenge for security and sovereignty.

A. National Awakening

1) The Emergence of a National Cyberdefence Authority

The development of the information systems, which are the nervous system of our societies, has been identified by France as a major vulnerability. As the *White Paper on Defence and National Security* [1] stated, *"information systems, which are the nerve system of our economic and social life, as well as of the operations of the public authorities, of the major energy, transport or food producers, or again the organisation of our defence, have made our societies and their defence vulnerable to accidental breakdowns or intentional attacks on computer networks."* All sectors of the nation are likely to be attacked, implying a brutal, deep and even durable destabilisation of the society: banking and financial systems, air and rail transportation networks, communication and media networks, energy and water production and distribution networks, state decision-making autonomy and governmental and military capacity of action. The security of these sectors has already organised against diverse threats, in particular terrorism, and has already imposed constraints on their public and private operators, called operators of critical infrastructures (OIV[2]). The French Defence Code states in its article L1332-1 that

> "[…] public or private operators which exploit some installations or use installations or facilities whose unavailability would seriously compromise the warfare or economic capabilities, the security or survivability of the nation, have to cooperate at their own expense […] in order to protect these installations, structures or facilities against any threat, particularly terrorism. These installations, structures or facilities are designated by the administrative authority."

These operators currently number more than 200 and are divided into seven sectors: state

1 1972, 1994, 2008 and probably after the national elections in 2012.
2 *Opérateur d'Importance Vitale* in French.

service; transportation; energy; health; communications; industry and finances; food and water management; space.

In addition to the daily massive attacks, generally poorly publicised in the media, many foreign examples have made the headlines: the paralysis of Estonia in 2007 showed the extreme vulnerability of digitised societies, while the war in Georgia in 2008 testified to the potential use of cyberspace in military operations.

According to the 2008 White Paper, the hypothesis of a large-scale IT[3] attack against national infrastructures is likely to happen in the next ten years:

> "Over the next 15 years, the proliferation of attempted attacks by non-State actors, computer pirates, activists or criminal organisations is a certainty. Some of these could be on a massive scale. With regard to attacks emanating from States, several countries have already mapped out offensive cyber-warfare strategies and are effectively putting in place technical capabilities with the aid of hackers. Covert attempted attacks are highly probable in this context. Massive overt actions are also plausible over the next fifteen years." [1].

The classic distinctions between state and non-state attack, as well as between the public or private status of the target, are blurred in cyberspace.

Drawing conclusions from this truly comprehensive, and not only military, nature of defence of the cyberspace, France created the French Network and Information Security Agency (ANSSI[4]) in 2009.

2) France Cyberstrategy

France has a long experience of inter-ministerial structures. Indeed, according to the Constitution[5], the Prime Minister is responsible for national defence. Under his direct authority, a Secretary General for Defence and National Security (SGDSN[6]) organises and coordinates all the ministries' policies relevant to this field. The ANSSI, which belongs to the SGDSN, saw its attributions enlarged in 2011: it is now the national authority for the defence of information systems. Thus, it has authority not only over the administration and public actors, but also over public and private operators of vital importance.

The ANSSI quickly proposed a national strategy [2] to give an orientation and to set priorities. This strategy is based on four objectives.

First of all, France must count among the top nations in the cyber effort in order to retain its strategic independence as well as cooperating at the highest level with other nations.
Then, France must guarantee its freedom of decision-making by protecting the information related to its sovereignty. Indeed, autonomy of decision and action supposes, in any situation, the confidentiality and availability of critical systems for information and communication. The indispensable security products, in particular cryptographic ones, must be nationally designed

3 Information Technology.
4 *Agence Nationale pour la Sécurité des Systèmes d'Information* in French.
5 Fifth Republic Constitution, 1958, article 21.
6 *Secrétariat pour la Défense et la Sécurité Nationale* in French.

or even produced.

Furthermore, considering French critical dependency on information and communication systems, especially on the Internet, every public and private actor must collaborate to guarantee the security and resilience of critical systems, in particular the equipments' producers and the operators of critical infrastructures.

Finally, beyond the control of cyberspace physical supports, security in this domain must be enforced. This task requires an important effort in the fight against criminality involving every actor: administrations, companies and citizens.

3) ANSSI Responsibilities

The ANSSI has a central role in this strategy. Responsible for the defence of information systems, its mission is to watch, detect, alert and react to computer and network attacks, in particular on governmental networks but also on the critical operators. In the case of a major IT attack against an administration or an operator of vital importance, the ANSSI can enforce defence measures, including the isolation of networks.

The ANSSI leads an operational centre for cyber defence (COSSI[7]) which is permanently watching sensitive networks and informs the CERTA[8] – the French governmental CERT. The ANSSI also assumes an important role in the conception, procurement and certification of trusted security products and services which are essential for the protection of the most sensitive networks[9]. It has elaborated a *Security General Framework*,[10] encompassing all the administrations.

ANSSI's growing power allows it to intervene in the most sensitive cases of cyber-incidents. It typically brought its assistance and savoir-faire into play in two very symbolic cases testifying to the high level of threat. In March 2011, more than 150 computers of the French Ministry of Economy, Finance and Industry were infected by a Trojan targeting documents about the G20 French Presidency. In September 2011, the French nuclear company Areva discovered a massive infection, which had lasted for more than two years and had potentially caused strategic damage.

B. The Case of the Ministry of Defence

1) The Specific Vulnerabilities of the Military Systems

Besides their instrumental information and communication role for the Ministry of Defence, the systems also condition the operational superiority of the armed forces:

> "information, as pointed out previously, is the key to all strategic functions […] In terms of operational military needs, in addition to the acquisition of information referred to

7 *Centre d'opération pour la sécurité des systèmes d'information* in French.
8 *Centre d'expertise gouvernemental de réponse et de traitement des attaques informatiques* in French.
9 Article 9 of the decree n°2005-1516, December, 8th, 2005.
10 *Référenciel Général de Sécurité* in French: set of rules drawn up by ANSSI and stipulated in Ordinance No. 2005-1516 of 8 December 2005 'on electronic exchanges between users and the public administration and between public administrations' that certain functions contributing to the security of information must comply with. This includes, among others, electronic signatures, authentication, confidentiality and times-tamps. The rules set out in the RGS are mandatory and are adjusted to reflect the level of security defined by the administrative authority concerning the security of the online services for which it is responsible.

under the 'knowledge and anticipation' strategic function, the object is to establish secure, reliable, protected and high capacity communications, from the highest level of the State down to those in the field." [1].

While these systems have to work at any time and, in particular, during exceptional circumstances, they face a double challenge. Designed for better interoperability and compatibility, many systems are based on Internet technologies, often designed without any security components. Thus, they can be victim of the very numerous widespread attacks on the Internet, the so-called 'background noise' of cyber incidents. For example, in 2009, the involuntary import of the Conficker Virus into the Navy network led to the temporary unavailability of this network while the virus was eradicated. But the logistical system of the Rafale combat aircraft, which is supported by that network, was compromised.

Furthermore, these military systems contain high-value information and contribute to the operational efficiency of the armed forces: they are specifically targeted by precise and tailored attacks, carefully planned and executed. These hostile actions can affect the systems and networks components of weapon systems: embedded systems, as well as the infrastructures or weapon platforms (including SCADA[11]).

2) The New Organisation of the Ministry of Defence

The Ministry has long experience in information systems security. But the increase of attacks and actors required a more proactive organisation, considering cyberspace as a new domain for warfare.

In July 2011, the Joint Concept for Cyberdefence [3] defined the objectives and principles of cyberspace control by the armed forces. The main goal is obviously to ensure an active and in-depth defence of information systems operated by the French armed forces for their homeland and overseas operations. But the Joint Concept also contributes to the continuity of the essential activities of the state and brings its support to French or foreign partners in the case of a major cyber crisis.

In January 2012, this concept was followed by a doctrine [4] aiming at organising the Ministry and creating an operational chain of command for cyberdefence. Broadly speaking, the Joint Chief of Defence Staff (CEMA[12]) is responsible for the employment and command of the armed forces. In cyberdefence, he is also in charge of the whole defence of the information systems of the Ministry. To lead this defence, a unique and centralised joint and ministerial chain of command is organised. A General Officer, directly connected with the Chief of Operations of the joint staff, is appointed to conduct the Defensive Cyber Operations (LID[13]) of the Ministry and to perform the executive management and coordination of the whole cyberdefence domain: organisation, human resources, procurement, etc.

This centralised organisation favours an exhaustive knowledge of cyber events and better coordination. The Joint Operations Planning and Command & Control Center (CPCO[14]) takes into account cyberdefence in military operations. Among the units dedicated to cyber, the

11 SCADA (Supervisory Control And Data Acquisition)
12 *Chef d'Etat Major des Armées* in French.
13 *Lutte Informatique Défensive* in French.
14 *Centre de planification et de conduite des opérations* in French.

Analysis Centre for Defensive Cyber Operations (CALID[15]) is in charge of surveillance of, analysis of and quick response to cyber attacks. True MOD-CERT, it is in close connection with the COSSI of the ANSSI (both centres will be colocalised in 2013) and it is the correspondent of the other allied military CERT.

3. INTERNATIONAL RELATIONS

A. Cooperation between States

The interconnection of networks, in particular via the Internet, raises questions around borders and principles of sovereignty. All modern states, including emerging countries, are now dependent on networks and, broadly speaking, suffer the same vulnerabilities.

Cybercriminals use the World Wide Web in order to commit trans-border crimes. By contrast, states have to manoeuvre to sue these criminals in a real, segmented world, where some countries do not recognise the illegality of cyber acts. For example, the French infraction of 'contestation of crime against humanity' ('Gayssot Act' of July 13th, 1990) is not recognised in most of the world's countries (in particular in the USA, in accordance with the First Amendment). Thus, a hacker can use a 'botnet' in order to block access to a website from different countries. By contrast, police investigators have to respect long multinational judicial cooperation processes. Public administrations and companies, as well as citizens, suffer the same vulnerabilities and the same attacks.

Thus, France is convinced of the added value of international cooperation to assure the best possible knowledge of emergent threats and to share solutions. To this end, the ANSSI, via the CERTA, establishes relations with its counterparts. Since September 2000, the CERTA is a member of the Forum of Incident Response and Security Teams (FIRST)[16] which includes more than 200 members, and takes part in the activity of the Computer Security Incident Response Team (TF-CSIRT)[17] (which is the coordination cell of the European CERT (Trusted Introducer Level 2 since March, 2002).

As a matter of fact, the CERTA is in touch with every country worldwide, except for a few countries in Africa and the Middle East which still lack the adapted structures.

1) NATO

The cyberdefence challenge was tackled at the Prague Summit in 2002. However, it was only stamped as a new official mission of the Alliance at the Lisbon Summit [5] in 2010. First of all, the cyberdefence policy aims at strengthening the NATO information system, thanks to the improvement of security standards and procedures, as well as a more centralised management. It was recognised a

> "necessity for NATO and the nations to protect the critical information systems according to their responsibilities, to share the best practices, to build up a capacity in order to assist, if required, the Alliance members to counter cyber attacks."

15 *Centre d'analyse en lutte informatique défensive* in French.
16 http://www.first.org/
17 https://www.trusted-introducer.org/index.html

Another objective is to strengthen NATO capacity to coordinate mutual assistance in case of an important cyber attack, possibly with projected teams.

The sharing of the burden between NATO and the nations, which are responsible for the protection of their own information systems, was defined in order to strictly delimit the perimeter of the systems to be shared. France, indeed, considers that the responsibility to protect national networks primarily lies with each ally.

The determination of a cyber action plan and the implementation of the adapted structures have happened particularly fast, testifying to the importance of the issue. The NATO Computer Incident Response Capability (NCIRC) should reach its full operational capacity as soon as possible. This equivalent of a CERT at NATO is the counterpart of the CALID, after the signature of a Memorandum of Understanding (MoU) between France and NATO in September 2011.

2) The European Union

Very early on, the European Union showed interest in new technologies. The European Commission initially considered cyber from the angle of the protection of critical infrastructures, as stated in many documents: the so called "i2010" strategy ("an information strategy for growth and employment", 2005), "Strategy for a secure information society", 2006, European Programme for Critical Infrastructures Protection (PEPIC), 2004 to 2007, Programme for crisis prevention, preparation and management in matter of terrorism and other security-related risks (CIPS), up to 2013.

But it still faces many hurdles. In spite of the adoption of the Lisbon Treaty in 2007, which would have led to a certain harmonisation thanks to the dissolution of the three pillars, the actors in charge of cyber issues are still numerous: six Directorates-General from the European Commission (DG Infso, DG Justice, DG Home, DG Entr, DG HR, DG JRC), General Secretary of the Council, EU External Action Service, Parliament, European Data Supervisor, European Network and Information Security Agency (ENISA), European Defence Agency (AED), Europol and the "common enterprises" (Galileo and Artemis; there is no common enterprise for information systems security itself). Moreover, those issues are dealt with separately, depending on the nature of the issue (protection of citizens, of economic or technological development, of critical infrastructures; fight against cybercrime; cyberdefence).

However, since 2004 the European Union benefits from a dedicated instrument within the European Agency in charge of networks and information security, the ENISA (European Network and Information Security Agency).

A unit for watch, alert and quick response at the disposal of European institutions (CERT-EU) should be entirely operational in May 2012, while the European IT agency for the area of freedom, security and justice, created on November, 1st, 2011, should be operational on December, 1st, 2012.

France widely supports these initiatives, which should increase security for the Member States and citizens of the Union. However, Paris regrets the lack of unity which hampers global

efficiency, and the absence of a military dimension, particularly critical in the case of any EU-led military operations. France also wishes to establish a stronger link between the EU and NATO, which have 22 members in common. The EU would take advantage of the advance of NATO in cyber, and would bring its own experience in civil crisis management.

B. World Governance

The transnational features of cyberspace make it a common space, just like space or the high seas. For now, the only binding international legal instrument managing relations between states in cyberspace is the Council of Europe Convention on Cybercrime, ("Convention of Budapest") [6]. This Convention was adopted in Budapest on November, 23rd, 2001, by the member states of the Council of Europe and their partners (USA, Japan, Canada, South Africa); it came into force on July 1st 2001. It was completed in 2003 by an Additional Protocol about racism and xenophobia via information systems. Up to now, 32 states have ratified this Convention. It imposes on the signatory states the obligation to set up a national legal framework necessary for the prosecution of crimes in and through cyberspace, and to set up judicial mechanisms of cooperation.

Other initiatives are beginning to blossom. On September, 12th, 2011, China, Russia, Uzbekistan and Tajikistan (members of the Shanghai Cooperation and Security Organisation) sent a "Code of conduct for information security" [7] to the General Secretary of United Nations, within the framework of the 66th General Assembly of the UNO. This code, insisting on the superiority of the national law in cyberspace, tries to legitimise a takeover of Internet governance by states in order to enforce their security in their 'informative spaces'. This proposal refers directly to a governance model which is more focused on contents (information) rather than on networks, considering information as a potential threat, and stressing the possibility for a government to challenge the political system of another state via the Internet. The initial intent of the submitting states was not to have this paper adopted during the General Assembly but to receive advice and comments, particularly from the perspective of the UN Group of government experts on information security, which will take place in August 2012.

Moreover, Russia considers the cooperation between States Parties as a legal form of espionage, and is dissatisfied with the condition of a consensus of all the Convention members for the admission of a state which is not a member of the Council of Europe. As a result, Russia followed up by proposing a "Convention on International Information Security" [8] in December 2011 during the international conference on security at Ekaterinburg.

These two 'information war' approaches raise obvious semantic issues. They oppose France and its Western partners, which consider governance in terms of 'information systems security', to the Chinese and Russian approach of 'information security', which could lead to an unacceptable censorship in cyberspace. For example, the project of a Code of Conduct equates the fight against terrorism with the fight against extremism and separatist activities.

Countries supporting these new proposals argue that there is a legal gap on the topic. They have not commented on the possible articulation of these proposals using the existing legal instruments. However, one can easily see a clear alternative to the Council of Europe Convention on Cybercrime, as far as these countries consider either the obsolete character of a ten-year-old

text (Chinese position), or the specific dimension of cyberspace which requires new rules in support of existing international law (Russian position).

Other initiatives have been launched in other *fora*, such as the ITU[18] or the OSCE[19]. But an initiative at the OSCE from the USA, which led the Cyber Steering Committee, would probably be rejected by China (a non OSCE member) and not supported by Russia; and the ITU, driven by its General Secretary Hamadoun Touré, wants to be involved in Internet regulation [9]. Its current orientation is not favourable to a universalisation of the Convention of Budapest and aims to support the Russian approach of "cyberarms" control. In consequence, they probably have less chance of success than a direct dialogue at the UNO, in particular through its Forum on the Governance of the Internet, the next meeting of which will take place in Geneva in February 2012.

However, the adoption of a resolution on cyberspace governance is still exclusively discussed within the First Committee of the UNO (Disarmament and International Security); this completely matches with the Sino-Russian proposals, and does not allow a more universal consideration of the cybersecurity issue. The meeting of the group of government experts (GGE) in August, 2012, where countries favourable to the Convention of Budapest will be a majority, but where Russia and China will have a blocking minority, constitutes an opportunity to discuss the Sino-Russian proposal and to reach a compromise. In contrast, a failure in this negotiation could fuel a logic of 'blocks', with numerous problems attached.

France's position is to support the Convention of Budapest, which offers a relatively loose framework for states and could contribute to the emergence of a consensus on a definition of the threat (cybercrime) recognised by all, even by the initiator states of the Code of Conduct. This base could then be enlarged to take into account the legitimate question of the nature of the information circulating on the Internet, related to personal data, intellectual properties, abuse of freedom of expression, paedopornography, etc., or international security issues.

4. CONCEPTS TO BE EXPLORED AND THE FUTURE OF CYBERSPACE

The surge in the use of information and communication systems is beginning to be seriously taken into account by numerous countries. However, many questions remain unsettled and new problems are appearing.

A. Public-Private Relationship

The private sector dominates cyberspace as the owner or the operator of most of the information and communication systems, as the designer and manufacturer of equipments, as the main user (through economic activity), etc.

1) Operators of Critical Infrastructures (OIV)

France has historically benefited from the legal instruments required to impose the necessary measures for the protection of critical infrastructures. It now needs to adapt them to the new challenges of cyberspace. A legal framework is necessary, but not sufficient: concrete and

18 International Telecommunication Union.
19 Organization for Security and Cooperation in Europe.

serious measures must be taken to ensure an effective security of the systems.

2) Security of Private Companies
Despite a general reduction of public jobs, the ANSSI staff is growing steadily, from 250 persons in 2012 to a target of 350 persons in 2013, particularly in order to perform its mission with private companies (even though it cannot guarantee the security of all the companies). To achieve those goals, a new organisation has been in place since April 2nd 2012.
That is why, in addition to legal measures and controls, the ANSSI also carries out advice and training. For instance, it promotes the concept of 'IT hygiene', which basically consists of implementing routine efficient security good practice, in particular, antivirus, passwords, security updates and appropriate administration procedures. The more complex technical and expensive solutions are only applied to counter targeted attacks.

3) Support of Private Sector
The role of the private sector is crucial in the development of the Defence Technological and Industrial Base (BITD[20]). As France wishes to maintain its ranking as a world-class country in security technologies, it has to set up tools enabling the private and public sectors to collaborate and improve their good practice together. This approach is gaining traction, but the shape it will take is not yet determined.

Beyond timely collaborations in the support for research and development as well as shared educational programmes, a promising path may be the creation of a hub gathering all the actors, based on the model of the cyber security hub proposed by the British cyber strategy [10].

B. Doctrine Issues

For defence, cyberspace is a source of new threats but also of opportunities. All the operation concepts have to be reviewed to integrate this new dimension and all the planning processes have to take it into account.

The rules of strategy and armed conflict are discovering a new field of application. As the *French White Paper on Defence and National Security* stated: *"as cyberspace has become a new action field in which military operations already take place, France has to develop a fighting capacity in this space."* [1]. The notions of "'cyberwar", "act of war", "dissuasion" have to be revisited, while the International Humanitarian Law and its principles (distinction between combatants and non-combatants, caution, proportionality, ban of unnecessary suffering) have to limit the use of cyberspace.

Last year, the French Defence University (IHEDN[21]), in partnership with EADS, created the "Castex Chair of Cyberstrategy" which stimulates high-level thinking on these concepts. At the level of the Ministry of Defence, studies are led by various institutions (Directorate for Strategic Affairs, Direction for Legal Affairs, Joint Centre for Concepts, Doctrine and Experiment) to take into account these new aspects of military action.

France also contributes to this thinking in international organisms such as NATO, and pays a

20 *Base Industrielle et Technologique de Défense* in French.
21 *Institut des Hautes Études de la Défense Nationale* in French, under the Prime Minister's authority.

close attention to the studies in ACT[22] and in the CCD COE[23].

C. The Future of Internet Governance

The properties of cyberspace call into question the concept of national sovereignty. Maybe John Perry Barlow went too far when he proclaimed the independence of cyberspace in 1996 [11]. Nevertheless, the traditional pillars of sovereignty face hurdles in mastering the dissemination of information streams.

1) Internal Sovereignty

Every state tries to control cyberspace, whether to guarantee the safety of its citizens (through the fight against cybercrime) or to enforce law and order (for instance, through censorship) On the one hand, the scope of the control depends on the openness of the regime. On the other hand, all states are confronted with the same technical and practical problems.

France views cyberspace as a neutral domain by default; only its use may deliberately cause damages and, as such, can be prosecuted. In particular, liberties as defined in the European Convention on Human Rights [12] have to be respected: freedom of thought, religion, expression, protection of privacy.

2) World Governance

The triangular relationship between states, companies – which are heavily present in cyberspace – and citizens – who use it massively – raises the issue of world governance striking a new balance in order to respect the rights and interests of every actor [13]. A promising framework for dialogue is the Internet Governance Forum, which allows real progress in international cooperation.

The lack of world regulation mechanisms, or the perceived illegitimacy of regulation itself, could fuel extreme behaviour from citizens ("Anonymous" is a famous example of the mode of action of "hacktivist" groups) and even lead to a sort of 'balkanisation' of the Internet, which would be segmented in regional networks and governed by different rules.

Although France is represented within the GAC (Governmental Advisory Committee) of the ICANN (Internet Corporation for Assigned Names and Numbers), it believes that the regulation of the Internet must be discussed and determined within the framework of the UNO and based on the principles of respect for individual freedoms.

5. CONCLUSION

In cyberspace as in other domains, France, which is a permanent member of the UNO Security Council and the fifth world economic power, wants to maintain its ranking. It has implemented a voluntarist policy to protect its critical infrastructures, to develop its security technologies and to integrate this new domain into military operations.

There are still considerable efforts to be made and this requires a real collective awareness on the part of all the actors: public and private sector and citizens.

France must also develop international cooperation agreements to share information about

22 NATO Allied Command for Transformation, Norfolk (USA).

23 NATO Cooperative Cyber Defence Centre of Excellence, Tallin (Estonia).

threats and solutions, as well as to promote the values of freedom and neutrality of the Internet. It is under this condition that 'the age of uncertainty or anxiety' [14] can become the age of prosperity and security.

REFERENCES

[1] *French White Paper on Defence and National Security*, 2008.
[2] ANSSI, *Information Systems Defence and Security: France's strategy*, February 2011.
[3] CICDE, *Concept interarmées de cyberdéfense* (CIA-6.3), July 2011.
[4] CICDE, *Doctrine interarmées de cyberdéfense* (DIA-6.3), January 2012.
[5] NATO, *Lisbon Summit Declaration*, November 2010.
[6] Council of Europe, *Convention on Cybercrime*, November, 23rd 2001.
[7] Shanghai Cooperation Organization, *Code of conduct for information security*, 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 (A/66/359).
[8] Russian Federation, *Convention on International Information Security (Concept)*, Ekaterinburg, Russia: International Meeting of High-Ranking Officials Responsible for Security Matters, 21-22 September 2011.
[9] Hamadoun I Toure, "The International Response to Cyberwar," in *The Quest for Cyber Peace*, International Telecommunication Union and World Federation of Scientists, January 2011.
[10] *The UK Cyber Security Strategy: Protecting and promoting the UK in a digital world*, November 2011.
[11] John Perry Barlow, *A Declaration of the Independence of Cyberspace*, Davos (Switzerland), February, 8th 1996.
[12] Council of Europe, *European Convention on Human Rights*, June 2010.
[13] Milton L. Mueller, *Networks and States: The Global Politics of Internet Governance (Information Revolution and Global Politics)*, September, 3rd, 2010.
[14] David J. Betz and Tim Stevens, *Cyberspace and the State: Towards a strategy for cyber-power*, 2011.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

A Treaty for Governing Cyber-Weapons: Potential Benefits and Practical Limitations

Louise Arimatsu
International Law Programme
Chatham House
London, UK
larimatsu@chathamhouse.org

Abstract: Despite a greater willingness on the part of States to enter into a dialogue on the potential implications of cyber warfare, there is continued disagreement on whether new rules are required to govern this 'new domain' and, if so, whether such rules should be in codified form or be left to evolve through a natural progression of customary international law. Closely interlinked with these questions is the distinct issue of whether there is a need for an arms control treaty. To speak of an arms control treaty or the regulation of a particular weapon by reference to the law of armed conflict (LOAC) is to presuppose a common conception of the particular type of weapon that is under discussion. This paper therefore poses the question, 'What is a cyber-weapon?' before considering whether an arms control treaty is a feasible option, let alone whether such a treaty would be capable of addressing the concerns that have been raised by its proponents. This paper also considers existing LOAC rules to identify the issues that are unique to cyber-weapons and, in doing so, it is argued that further clarification is indeed merited.

Keywords: *cyber-weapons, arms treaty, law of armed conflict*

1. INTRODUCTION

On 12 September 2011 China and the Russian Federation, together with Tajikistan and Uzbekistan, submitted a draft United Nations General Assembly resolution on an *International code of conduct for information security.*[1] The unexpected move, just prior to a global conference on cyberspace, was described by some as an attempt to 'regain the initiative' on a topic that has commanded increasing attention by the international community over the last several years.[2] The draft code requires States to comply with the UN Charter and 'universally recognized norms governing international relations that enshrine, inter alia, respect for the sovereignty, territorial integrity and political independence of all States' and 'not to use information and

1 UN GA Doc. A/66/359 of 14 September 2011.
2 For details on 2011 London Conference on Cyberspace see http://www.fco.gov.uk/en/global-issues/london-conference-cyberspace/.

communications technologies, including networks, to carry out hostile activities or acts of aggression, pose threats to international peace and security or proliferate information weapons or related technologies'. Though described as a draft voluntary code, this document not only illuminates what interests are perceived to be at stake by the sponsoring States but exposes the potential difficulties that would be encountered in any attempt to negotiate a cyber treaty, not least one that is concerned with stemming the proliferation of cyber-weapons or what the draft code refers to as 'information weapons'.

The potential benefits and practical limitations of a treaty to govern cyber-weapons can only be fully appreciated with an understanding of the historical and political context within which the cyber discourse has evolved in recent years. This paper therefore opens with a brief look at the context and the issues upon which States have traditionally divided to assess whether there is any prospect for agreement (section 2). While such divisions are founded primarily on disparate ideological and political views on the role of the State, legal experts also differ on whether new rules are required to govern cyber warfare and, if so, whether such rules should be in codified form or be left to evolve through a natural progression of customary international law. Often intermingled with this question is the distinct issue of whether there is a need for an arms control treaty of sorts, as inferred by the Sino-Russia draft code of conduct. The objective, according to the proponents of such a treaty, is to limit the digital or cyber 'arms race' between States, with a view to constraining or even prohibiting the use of cyber-weapons in certain circumstances.[3] Obviously these questions are not unique to the cyber warfare discourse. Progress in the realms of science and technology, which invariably feeds into warfare, has always prompted similar anxieties.

In section 3 of the paper, I pose the simple, yet often ignored question, *what is a cyber-weapon*? I do so because to speak of an arms control treaty or the regulation of a particular weapon by reference to the law of armed conflict (LOAC) is to presuppose a common conception of the particular type of weapon that is under discussion. And although the term 'cyber-weapon' is entrenched throughout the policy and legal literature on cyber warfare, it is telling that in November 2011, the US Department of Defense stated, '[t]here is currently no international consensus regarding the definition of "cyber weapon"'.[4] From this, should we surmise that there is something unique about the cyber-weapon that inhibits definitional agreement? What are the attributes that distinguish such weapons from conventional weapons and do these tell us anything about why agreement continues to prove elusive?

In this context I consider whether an arms control treaty is a feasible option let alone whether such a treaty would be capable of addressing the concerns that have been raised in respect of the prospect of cyber warfare.[5] International law has historically dealt with weapons through two parallel approaches: regulating the manner in which weapons are used or by focusing on a particular type of weapon.[6] Whether cyber-weapons are better suited to be governed by LOAC

3 John Markoff and Andrew Kramer, 'U.S. and Russia Differ on a Treaty for Cyberspace' 28 June 2009, New York Times. Franz-Stefan Gady and Greg Austin, 'Russia, the United States and Cyber Diplomacy', EastWest Institute paper 2010, 6.

4 United States Department of Defense Cyberspace Policy Report: A Report to Congress Pursuant to the National Defense Authorization Act for Fiscal Year 2011, Section 934, November 2011, 8.

5 The problem of attribution is probably of most concern.

6 Justin McClelland, 'The review of weapons in accordance with Article 36 of Additional Protocol I' IRRC (2003) Volume 85, No. 850, 397.

or whether an arms control treaty is warranted will be explored given the different rationale upon which each approach is founded. In the penultimate section I ask whether, all things considered, existing LOAC rules are adequate but that, nonetheless, the particular context of cyber warfare demands greater clarity as to how the rules are interpreted and applied and what form this might take.[7] Not all experts share the view that the law in its current form can respond fully to the particularities of the cyber challenge; as a consequence, some have called for far more proactive measures including a treaty to govern cyber warfare.[8] The recent developments at the international level would suggest that there may be an emerging consensus among States in favour of a set of agreed rules governing cyber warfare more generally although both form and content may be difficult to secure.

I conclude with some thoughts on areas for further exploration.

Before proceeding, one note of caution is required. A persistent problem that has characterised this entire discourse is the prevalence of misleading language and 'parallel vocabularies' in discussions on all aspects of cyber space and security.[9] For example, as one legal expert has noted, despite widespread use of the terms 'cyber warfare' and 'cyber attack', the vast majority of cyber activity targeting the U.S. cannot, under existing law of armed conflict, be described as an 'attack' that would give rise to a situation of armed conflict operationalising that body of law.[10] This problem is compounded by the strategic choice of some of the leading players to use phrases that are broad in scope to safeguard what are genuinely regarded as legitimate sovereign interests, made even more pressing by the extent to which the Arab Spring revolutions were facilitated by the digital revolution.

7 C. Joyner and C. Lotrionte, 'Information Warfare as International Coercion: Elements of a Legal Framework' 12 EJIL (2001) 825-65.

8 Davis Brown, 'A Proposal for an International Convention to Regulate the Use of Information Systems in Armed Conflict' 47 Harvard International Law Journal, 179-221.

9 Russia-U.S. Bilateral on Cybersecurity - Critical Terminology Foundations, K. F Rauscher and V. Yaschenko (eds), 2011 EastWest Insitute and the Information Security Institute of Moscow State University. See also 'Technology, Policy, Law and Ethics Regarding U.S. Acquisition of Use of Cyberattack Capabilities' W. Owens, K. Dam and H. Lin (eds) National Research Council (2009) 14-15, Box1.2 [hereinafter Technology, Policy, Law]; according to the report, under current US military doctrine computer network operations include computer network attack (CNA), computer network defense (CND) and computer network exploitation (CNE), 161.

10 Commander Todd C. Huntley, 'Controlling the use of force in cyber space: the application of the law of armed conflict during a time of fundamental change in the nature of warfare' 60 Naval Law Review (2010) 2.

2. THE CONTEXT

The United States' decision in October 2009 not to oppose a draft UN General Assembly resolution to explore possible measures to 'strengthen information security at the global level' signalled a fundamental shift in its cyber security policy.[11] For over a decade the US had resisted the repeated attempts by Russia – under the auspices of the UN Committee on Disarmament and International Security[12] (hereinafter First Committee) – to explore the possibility for formalising the rules pertaining to cyber security. Ideological differences coupled with mistrust as to motive on the part of *both* sides had created gridlock and it was only with significant redrafting of Russia's 1998 draft resolution to address US concerns combined with the re-assessment by the Obama administration in 2009 that US cyber strategy would benefit from greater international engagement, that progress, albeit limited, was secured.[13]

These developments paved the way for the release, in July 2010, of a report by the Group of Governmental Experts (GGE) comprising cyber security specialists and diplomats representing 15 countries including Russia and the US.[14] By contrast to an earlier attempt in 2005 the GGE, established by General Assembly resolution 60/45, was able to reach agreement in respect of a number of recommendations.[15] These included: to pursue further dialogue among States to discuss norms pertaining to the use of information and communication technologies (ICTs); to consider measures to address the implications of ICTs by States in situations of armed conflict; and to explore possibilities for elaborating on common terms and definitions.[16] Although the challenges identified in the GGE report are of pressing concern to all States, profound disagreements founded on radically differing perspectives and perceived interests are likely to hinder speedy progress, at least insofar as any treaty regime is concerned.

Russia and the US have approached the issue of cyber security from fundamentally different legal perspectives with the former favouring the development of a binding international regime while the latter has treated cyber security as falling, first and foremost, within a law enforcement paradigm and therefore better governed through suppression conventions and

11 Draft Resolution A/C.1/64/L.39 (16 October 2009) on Developments in the field of information and telecommunications in the context of international security was adopted by the General Assembly without a vote on 29 October 2009; see http://daccess-dds-ny.un.org/doc/UNDOC/LTD/N09/563/73/PDF/N0956373.pdf?OpenElement and http://www.un.org/News/Press/docs/2009/ga10898.doc.htm.

12 The Disarmament and International Security Committee deals with disarmament and related international security questions.

13 Recognizing that the US could not work in isolation if it wanted to succeed in cyberspace, the review called for a 'strategy for cybersecurity designed to shape the international environment and bring like-minded nations together on a host of issued such as technical standards acceptable legal norms regarding territorial jurisdiction, sovereign responsibility, and use of force.' The review was released in May 2009 and can be found at http://www.whitehouse.gov/assets/documents/Cyberspace_Policy_Review_final.pdf.

14 A/65/201 of 30 July 2010. The GGE was established under UN GA resolution 60/45.

15 On 11 January 2011, the General Assembly welcomed the report and took note of the recommendations contained therein; a new paragraph was introduced requesting the Secretary-General to establish a new GGE in 2012 to submit a report at the 68th session in 2013 (A/RES/65/41). See also A/RES/66/24 of 13 December 2011 adopting draft resolution A/66/407, 10 November 2011.

16 Report by the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, 30 July 2010, A/65/201.

mutual assistance.[17] While both regard cyberspace as a domain of economic opportunity as well as of heightened risk, for the US, the primary threat is criminal rather than political in origin.[18] Consequently, it regards Russia's demands for a broad international legal regime as overly prescriptive and fears that any concessions made on its part will assist in legitimising State censorship and repressive domestic policies. Such concerns are not without foundation. The very term 'information security' preferred by Russia, and often equated to 'cyber security', belies the reality that it is a far more 'sweeping concept tied to the State's need for control over the information space of its citizenry'.[19] In light of the *Information Security Doctrine of the Russian Federation* released by President Putin in 2000, it is difficult to escape the impression that Russia's broader concern is with how it can effectively maintain social control of the Internet in the face of both external and internal challenges.[20] At its most basic, the different approaches pursued are primarily, although not exclusively, a reflection of the different ideological viewpoints on the role of the State.[21]

A supplementary reason driving Russia's ambitions for an international cyber arms control treaty (and one that must not be under-estimated) is its perceived inferiority in the field of communications technology.[22] Although the US's investment in, and reliance on, information technology – whether civilian or military – may in the short term make it far more vulnerable to malicious digital intrusions, Russia's reliance on commercial off-the-shelf hardware and

17 The US is party to the Convention on Cybercrime (Budapest Convention) which it ratified in September 2006. The Convention was drafted by the Council of Europe (COE) and despite its official 'observer' status, the US played an 'especially influential role, in part because it had more experience than other countries in addressing cybercrime and entered the process with well-formulated positions'; Michael Vatis 'The Council of Europe Convention on Cybercrime' in Proceedings of a Workshop on Deterring CyberAttacks: Informing Strategies and Developing Options for U.S. Policy, National Academies Press (2010) 207, available at http://www.nap.edu/catalog.php?record_id=12997. By contrast, although Russia is a member of the COE, it has neither ratified nor signed the Convention on the grounds that it views the provision allowing unilateral trans-border access by law enforcement agencies to computers or data with the consent of the computer- or data-owner, as a violation of sovereignty; Vatis 218.

18 As a UN report acknowledged in 2010, it is difficult to estimate the extent of the financial loss and number of offences committed by cybercriminals. Although guarded, the report refers to some sources estimating losses to businesses and institutions in the US due to cybercrime to be worth as much as US$67 billion per year; A/CONF.213/9 of 22 January 2010. According to the 2011 Norton Cybercrime Report the cost of global cybercrime stands at US$114 billion annually.

19 Christopher A. Ford, 'The Trouble with Cyber Arms Control' The New Atlantis 2010, 52-68, 63. See also Timothy L. Thomas, 'The Russian Understanding of Information Operations and Information Warfare' in Information Age Anthology: the Information Age Military, D. Alberts and D. Papp (eds) 2001, available at www.dodccrp.org.

20 The Information Security Doctrine of the Russian Federation, 9 September 2000, available at http://www.mid.ru/ns-osndoc.nsf/osnddeng.

21 This observation might apply equally to China which also adopts a far broader understanding of cyber threats; Ford, 'The Trouble with Cyber Arms Control', 62-66. That Russia and China share many of the perceived threats is best exemplified by the Agreement between the Governments of the Member States of the Shanghai Cooperation Organization on Cooperation in the Field of International Information Security which was adopted at the 61st plenary meeting of the Organization on 2 December 2008. Nevertheless, there is also evidence to indicate that at a strategic/operational level, China may adopt a radically different approach from Russia in that it sees cyber warfare as an 'equalizer' in potential military conflicts with a technologically superior adversary such as the US; Technology, Policy, Law, 332-33.

22 Information Security Doctrine of the Russian Federation.

software and lack of home-grown expertise makes it far more vulnerable in the long run.[23] Thus, as in 1899, Russia envisages that an international treaty may function to address its position of relative disadvantage.[24] While much of the cyber security discourse over the last decade has been dominated by Russia and the US, China's emerging status and participation within this field poses questions that have yet to be explored adequately. How it frames both opportunities and risks in this domain is likely to shape any global progress on an internationally agreed regime.[25] Finally, it is necessary to ask whether, despite the US's long held scepticism over the prospect of an arms control treaty, there are any emerging or potential future benefits or threats that might alter its stance in favour of such a treaty.

Since taking office, the Obama Administration has responded robustly to its critics' charges that a comprehensive national security strategy that embraced both the domestic and international aspects of cyber security was lacking.[26] In May 2011, the White House released its *International Strategy for Cyberspace* – 'the first attempt by the US to lay out an approach that unifies its engagement with international partners on the full range of cyber issues' and it concurrently embarked on a variety of outreach efforts to enhance existing military alliances and to pursue closer cyber security partnerships with like-minded States. But despite its willingness to engage in a more pro-active dialogue, what is striking is the Administration's continued emphasis on developing the law enforcement paradigm to counter malicious cyber activities with little evidence to suggest that there has been a fundamental shift in its position on the need for an arms control treaty. Rather, the message that is repeatedly heard is that existing international law suffices.[27]

Any consideration as to whether a new treaty regime to govern a particular weapon is necessary

23 'Russia's international cooperation in ensuring information security has two distinctive features: international competition for technological and information resources and for dominance in the markets has increased, and the world's leading economies have achieved a growing technological lead that allow them to build up their potential for information warfare. Russia views this development with concern, as it could lead to a new arms race in the information sphere and raises the threat of foreign intelligence services penetrating Russia through technical means, such as a global information infrastructure.' A.A. Streltsov, State Information Policy: The Basis of the Theory, 2010, Moscow, 345 cited by Franz-Stefan Gady and Greg Austin, 'Russia, the United States and Cyber Diplomacy' EastWest Institute (2010) 6.

24 According to Jozef Goldblat, 'the Hague Conferences of 1899 and 1907 were convened at the initiative of the Emperor of Russia, which was lagging in the European arms race and could not afford to catch up with its rivals because of its economic weakness'; Arms Control: The New Guide to Negotiations and Agreements (2003) Sage Publications, section 2.1.

25 See for example, submissions by China's representative to the 17th meeting of the First Committee, 20 October 2011, GA/DIS/3442.

26 James A. Lewis, 'Cyberwarfare and its Impact on International Security', United Nations Office for Disarmament Affairs (UNDOC) Occasional Paper No. 19 June 2010; 'Securing Cyberspace for the 44th Presidency: A Report of the Center for Strategic and International Studies Commission on Cybersecurity, December 2008. At the domestic level, U.S. Cyber Command (USCYBERCOM) was established in June 2009 with the responsibility for centralizing command of cyberspace operations. Nevertheless, see also US Government Accountability Office (GAO) report, Cyberspace: US Faces Challenges in Addressing Global Cybersecurity and Governance (GAO-10-606) July 2010.

27 See for example the EU-US Working Group on Cybersecurity which was established in November 2010 and tasked principally with strengthening transatlantic cooperation in the field of cyber-crime, PRES/10/315 available at http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/246. The most recent Department of Defense reports indicate that the US remains unconvinced that a convention governing cyberwar or more specifically, cyber-weapons is warranted; see for example, Department of Defense Strategy for Operating in Cyberspace, July 2011 and Department of Defense Cyberspace Policy Report to Congress Pursuant to the National Defense Authorization Act for Fiscal Year 2011, Section 934 of November 2011.

and, if so, what form this might take requires an understanding of the nature of the weapon under examination. In the following section I pose the simple yet vexing question, '*what is a cyber-weapon?*' before assessing whether an arms control treaty will adequately address the concerns raised by its proponents.

3. AN ARMS CONTROL TREATY FOR CYBER-WEAPONS

By contrast to most commentaries on cyber warfare where the preliminary concern is with the question what is an *'armed attack'*,[28] this paper asks, *what is a cyber-weapon*? A weapon is generally understood to be an instrument of offensive or defensive combat and has been defined as a device that is 'designed to kill, injure, or disable people, or to damage or destroy property''.[29] Although this definition might adequately encapsulate traditional weapons that have been designed, when utilized, to have a direct kinetic outcome, it fails to capture the essence of what are generally regarded as cyber-weapons. This is because most of the malicious codes or malware that would fall within the parameters of a cyber-weapon are designed to have an *indirect* kinetic outcome which may, or may not, result in the listed outcomes. In other words, the malware itself is not designed to kill, injure or disable people nor, necessarily, to damage or destroy tangible property. Moreover, even if 'property' is to encompass digital network systems, programmes and data, this particular definition is arguably under-inclusive if 'damage' or 'destruction' of property is narrowly defined. For example, the purpose of Duqu, a remote access Trojan which was discovered in September 2011 and believed to have been invented by the same authors as Stuxnet, was to *gather intelligence data and assets to enable an attack* by a worm such as Stuxnet.[30] Duqu was designed neither to damage nor destroy, yet there is evidence to suggest that the capacity of Stuxnet to achieve *its* design objective was dependent on the prior implanting of Duqu, which went undetected for four years. Nonetheless, the suggestion that a 'cyber-weapon' might be defined by its capacity for inflicting 'harm' is unconvincing for being over-inclusive.[31]

An alternative definition that begins to address the shortcomings of the above definition is any 'malicious software that possesses an offensive capability'.[32] The problem with this definition is self-evident. As with the term 'cyber-attack' which is commonly used to describe any action ranging from penetrating a network and implanting malicious codes, to downloading information and disrupting the services provided by those networks, it lacks the specificity that

28 Technology, Policy, Law, 1-2.

29 G. Intoccia and J. Wesley Moore, 'Communications Technology, Warfare, and the Law: Is the Network a Weapon System?' 28 Houston Journal of International Law (2006) 467-489, 480 citing Air Force guidance can be found in AFPD 51-4, which addresses Air Force regulatory compliance with LOAC and defines.

30 Brigid Grauman, 'Cyber-security: the vexed question of global rules' Security and Defence Agenda Report, February 2012, 30. See also Symantec Security Response briefing paper 'W32.Duqu: The precursor to the next Stuxnet' 23 November 2011; available at http://www.symantec.com/connect/w32_duqu_precursor_next_stuxnet.

31 Dorothy Denning, 'Reflections of Cyberweapons Controls' 16(4) Computer Security Journal (2000) 43-53, 46.

32 This definition borrows from one that was used to define conventional weapons as having 'an offensive capability that can be applied to a military object or enemy combatant'; McClelland, IRRC (2003) Volume 85, No. 850, 397-415, 405.

is necessary for legal regulation.[33] Thus, just as it is now generally recognised that it is the *effect* of the cyber-attack that determines whether or not the law of armed conflict is operationalised, a malicious code might be deemed a 'weapon' not solely by its intrinsic properties but also by the outcome it is designed to produce.[34] In other words, only if it is established that a malicious code possesses an offensive capability and there is an intention to use it in a manner which comports with its offensive capability might the malware be deemed a 'cyber-weapon'. Accordingly, it is both the offensive capability of the malicious code and the intended outcome or effect produced by that code that transforms it into a weapon that would be governed, as with any conventional weapon, by the law of armed conflict.[35]

In June 2011, it was reported that the Pentagon had developed a classified list of cyber-weapons and cyber-tools including viruses with the capacity to sabotage an adversary's critical networks.[36] This announcement would seem to suggest that the absence of international consensus on a definition for a 'cyber-weapon' might be indicative of a political impasse rather than there being any intrinsic attribute that precludes cyber-weapons from definition.[37] However, there is an enormous gulf between policy assessments and legal classification and since most of the technology relied on in an offensive capacity is inherently dual-use, and non-malicious 'software might be minimally repurposed for malicious action', drawing the line between the two is likely to be hugely challenging.[38] Even if agreement can be reached on what constitutes a cyber-weapon, whether their very properties make cyber-weapons simply incompatible with the rationale upon which arms control treaties are founded is warrants consideration.

Broadly stated, arms control treaties aim to establish legal regimes that 'deter challenges to peace'.[39] There are various categories of arms control and disarmament treaties that can broadly

33 James A. Lewis, 'Cyberwarfare and its impact on international security' (2009) UNODA Occasional Paper No. 19, 8.

34 Michael Schmitt, 'Cyber Operations and the Jus in Bello: Key Issues' 87 International Law Studies, Naval War College (2011); Charles Dunlap 'Perspective for Cyber Strategists on Law for Cyberwar' Strategic Studies Quarterly (Spring 2011) 81-99, 85.

35 This however does not resolve the definitional problem in its entirety since there will be some digital tools that only if directed at or used in a certain manner will produce an outcome, albeit indirectly, that can be equated to other traditional weapons. Since the same 'cyber-weapon' deployed in a different manner may result in an effect that is simply disruptive, regulating the use of the weapon, rather than the weapon per se may present a more viable option.

36 Ellen Nakashima, 'List of cyber-weapons developed by Pentagon to streamline computer warfare' in The Washington Post, 1 June 2011.

37 In the Shanghai Cooperation Organization's agreement on Cooperation in the Field of International Information Security which was adopted at the 61st plenary meeting of the Organization on 2 December 2008 'information weapon' is defined very broadly as 'information technologies, ways and means of waging an information war'.

38 United States Department of Defense Cyberspace Policy Report: A Report to Congress Pursuant to the National Defense Authorization Act for Fiscal Year 2011, Section 934, November 2011, 8.

39 Richard Betts, 'Systems for Peace or Causes of War? Collective Security, Arms Control and the New Europe' 17 International Security (1992) 5-43, 30. It is worth pondering on whether the absence of a cyber-weapons control convention has the counter-intuitive effect of promoting deterrence. The risk that an adversary has already developed and implanted malware that has the capacity to control a State's offensive and defensive capabilities may in fact serve to deter the kinetic use of force. That a State may not realise until some time later that its military capabilities have been eroded, may have a beneficial effect of instilling caution.

be grouped[40] into those that: i) limit the level of armaments;[41] ii) prohibit or restrict the use of specific weapons;[42] iii) prohibit the testing and deployment and attacks on the environment;[43] and iv) prohibit development and acquisition of specific weapons.[44] By contrast to the law of armed conflict, the objective of such regimes is to make conflict less likely by reducing the existence of, or restricting the use of certain weapons irrespective of whether the particular weapon is more or less cruel or indiscriminate than others which may not be the subject of such negotiations.[45] In addition to reducing the risk of armed conflict by imposing limitations on the development and proliferation of weapons to constrain capabilities, the purpose of such regimes can include:

- minimizing disparities among States to remove the source of instability;
- increasing predictability in relations between potentially hostile States;
- pre-empting the development of new weapons;
- decreasing expenditure on armaments to divert resources to economic and social development;
- contributing to conflict management by establishing a framework to enable negotiation between belligerent States;
- generally fostering a non-hostile atmosphere; and
- alleviating the suffering and damage in armed conflict.[46]

The distinction between the objectives of an arms treaty and LOAC is worth noting since cyber-weapons do not directly inflict the harm that LOAC is concerned with regulating. Arms treaties by contrast are generally agreed to not on the basis that the weapon is, all things considered, offensive to fundamental LOAC principles but rather because, as a matter of military and political judgment, the new restrictions can be the subject of agreement.[47] The agreement is treated as a 'contractual undertaking' adopted on the basis of a common interest: in other words, arms control treaties are the product of a policy choice rather than a legal necessity.

40 These categories of agreements were identified by Frits Kalshoven in The Centennial of the First International Peace Conference: Reports and Conclusions (2000) Kluwer Law International, 61-96.

41 1990 Treaty on Conventional Armed Force in Europe; 1972 Strategic Arms Limitation Talks I (SALT I); 1972 Anti-Ballistic Missile Systems Treaty; 1979 SALT II; the 1987 Intermediate-Range Nuclear Forces Treaty; 1991 Treaty on the Reduction and Limitation of Strategic Offensive Arms (START I); 1993 START II; START III.

42 1925 Geneva Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or Other Gases and of Bacteriological Methods of Warfare; 1981 Convention on Excessively Injurious or Indiscriminate Conventional Weapons and Protocols.

43 1963 Partial Test Ban Treaty; 1974 Threshold Test Ban Treaty; 1996 Comprehensive Nuclear Test Ban Treaty; 1977 Convention on the Prohibition of Military or Any Other Hostile use of Environment Modification Techniques; 1959 Antarctic Treaty; 1967 Outer Space Treaty; 1971 Seabed Treaty.

44 1968 Non-Proliferation Treaty; 1972 Biological Weapons Convention; 1993 Chemical Weapons Convention; 1997 Ottawa Convention on Anti-Personnel Mines.

45 Christopher Greenwood, 'The Law of Weaponry at the Start of the New Millennium' in Essays on War in International Law (2006) Cameron May, 223, 231.

46 Daniel Frei 'International Humanitarian Law and Arms Control' IRRC, No. 267 November-December 1988, 491, 493-94. According to Goldblat, compared to its original narrow meaning to denote rules for limiting arms competition, 'arms control' is now often used to refer to a broad range of measures including those intended to: freeze, limit, reduce or abolish certain categories of weapons; ban the testing of certain weapons; prevent certain military activities; regulate the deployment of armed forces; proscribe transfers of some military items; reduce the risk of accidental war; constrain or prohibit the use of certain weapons or methods of war; and build up confidence among States through greater openness in military matters; Goldblat, Arms Control: The New Guide to Negotiations and Agreements, 3.

47 Ashley Roach, 'Certain Conventional Weapons Convention: Arms Control or Humanitarian Law?' 105 Military Law Review, (1984) 3-72, 17.

What is of note is that such treaties have usually aimed to construct a military balance between States based on the simple reasoning that a parity in available arsenal would in itself dissuade the resort to force because it cannot be effectively exercised.[48] Thus, the key to such treaties is the ability to maintain a balance of power between *States*.[49] This is a perfectly reasonable rationale if the particular weapon is predominantly accessible – and affordable – only to States, as in the case of nuclear weapons. In the case of malware, this rationale offers little traction. Compared with other kinetic weapons, malicious software is easy to use and relatively cheap. These two factors make cyber-weapons widely accessible to non-state actors – from criminal gangs to the lone hackers. According to McAfee, every year sees one million new viruses, from worms to logic bombs; and that figure is climbing.[50] Moreover, unlike other weapons, cyber-weapons can be reproduced and distributed globally at minimal cost.[51] Even if a significant proportion of these malicious codes are generated by State actors, that still leaves a large number being created in the private sector. In the face of the sheer volume at which malware is being constituted, particularly by non-state actors, demands for an arms treaty comparable to the Chemical Weapons Convention (CWC)[52] to prohibit 'the development, spread and use of the 'information weapon'' appears a daunting, if not futile, exercise.[53]

Nevertheless, would it be feasible to introduce a system of classification for cyber-weapons linked to the level of harm that could potentially be caused by the malware? In other words, to adopt an approach comparable to the CWC and to focus efforts on malicious codes which have been designed primarily with offensive capabilities, the use of which is likely to result in serious harm comparable to a kinetic weapon? The CWC may also offer guidance in respect of exclusion clauses, as for example, malware that is produced for the very purpose of enabling the development of new programmes to detect and counter the intended harm. But what it cannot do is to provide a template. The speed at which technology is evolving means that the methods and tools of attack are constantly altering making any listing of prohibited cyber-weapons simply redundant.

For the purpose of argument, if distinguishing between offensive and defensive cyber-weapons is possible and the former is made subject to prohibitions, this still leaves the problem of dual-use software. As the DoD noted in its 2011 report to Congress, 'most of the technology used in this context is inherently dual-use, and even software might be minimally repurposed for malicious action.'[54] Although the issue of dual-use arose in the case of the CWC and the Nuclear Non-Proliferation Treaty (clearly chemical products and nuclear energy can be produced for

48 Betts 'Systems for Peace or Causes of War?' 30.

49 This reasoning was based primarily in the context of a bipolar world in the context of nuclear weapons. Experts have suggested that where there is more than just one pair of competing powers with overlapping rivalries, arms races are likely to be interconnected, and the stability of any one pair of rivals might be affected negatively by developments in other dyads. This means that there is even greater risk of instability and this 'increased political complexity of the post-bipolar world calls for more rather than less arms control.' Harald Muller 'Compliance Politics. A Critical Analysis of Multilateral Arms Control Treaty Enforcement'The Nonproliferation Review (2000) 77-90, 78.

50 Grauman, 'Cyber-security: the vexed question of global rules', 10.

51 In addition, in contrast to for example chemical weapons, cyber-weapons can be stored with no physical risk.

52 J. Markoff and A. Kramer, 'U.S. and Russia Differ on a Treaty for Cyberspace' New York Times, 28 June 2009.

53 See 2000 Russian Federation Information Security Doctrine, section 7 on 'International cooperation by the Russian Federation in the realm of information security'.

54 United States Department of Defense Cyberspace Policy Report: A Report to Congress Pursuant to the National Defense Authorization Act for Fiscal Year 2011, Section 934, November 2011, 8.

both peaceful and non-peaceful purposes) in the case of cyber-weapons, the question of dual-use takes on another dimension. This is because cyber-weapons possess multiple properties (e.g. destroy, degrade, exploit, control, deceive, alter) and thus to describe them as '*dual*-use' is potentially misleading.[55] The distinction between malware that seeks to exploit (e.g. for commercial gain) or is implanted to gather intelligence (e.g. State-sponsored espionage) and malware that is potentially offensive (to destroy or control) is tenuous at best as the Duqu/Stuxnet example aptly demonstrates.

Of course, irrespective of motive, the intruder must first be able to access a system or network and identify vulnerabilities in the hardware, software, hardware-software interfaces, communication channels, configuration tables, users, and/or service providers.[56] But the 'payload' or the malicious code or programme that performs a particular action, once a vulnerability has been detected, can take many forms. A bot or botnet is sometimes designed to disable websites and networks and sometimes to gather information,[57] a 'logic bomb' which is hidden in computers to halt them at crucial times or damage circuitry is designed to degrade or destroy, a microwave radiation device that can burn out computer circuits from a distance is principally designed to destroy, a distributed denial of service (DDoS) programme aims to disrupt, and other hacking tools including viruses, worms, spyware, or Trojan horses can be designed to perform one or a combination of operations. This attribute of cyber-weapons means that, rather than identifying specific categories of malware that would be subject to prohibition, it would seem far more effective to regulate the use of such weapons.

A further distinguishing property of malware is that in contrast to conventional weapons where the State has full control over the means by which weapons are deployed, it is the private sector or individuals who have ownership and operational rights over networks.[58] As a consequence any treaty system would require, at a minimum, a commitment on the part of the private sector to collaborate in what will likely be an operation of unprecedented complexity.

What will however be the Achilles' heel of a cyber-weapons control treaty is non-compliance since there is little prospect of integrating a reliable verification mechanism into such a treaty regime. It is unlikely that any State would agree to external verification measures which would necessarily require scanning all computers and storage devices owned and used by the State including all classified systems.[59] This is a significant drawback as past experience demonstrates that the success of arms control treaties has been contingent largely on the existence of a robust compliance and verification regime. For example, although the Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on Their Destruction (BWC) entered into force in 1975 and has 165 State Parties, it has repeatedly been criticised for lacking credibility on the grounds that it contains

55 Denning has described 'dual use' weapons to include password crackers and vulnerability and port scanners; as she notes, 'great caution would be required as many of these tools help system administrators find and correct security problems'; Dorothy Denning, 'Reflections on Cyberweapons Control' 16(4) Computer Security Journal (2000) 43-53, 43.

56 P. Denning and D. Denning, 'Discussing Cyber Attack' 53(9) Communications of the ACM (2010) 29.

57 Botnets may be designed simply to gather information; botnets refer typically to ordinary computers hijacked by viruses to perform attacks without their owner's knowledge; Duncan Hollis 'Why States need an international law for information operations' 11 Lewis & Clark Law Review (2007) 1023-1061, 1025.

58 GGE report, 6, A/65/201.

59 Dorothy Denning, 'Obstacles and Options for Cyber Arms Controls' presented at Arms Control in Cyberspace, Heinrich Böll Foundation, Berlin, Germany, 29-30 June 2001, 3.

no effective verification provisions.[60] The combined effect of having no means by which to independently verify compliance together with the ease at which malware can be secreted and the high degree of anonymity in cyber-space which makes the tracking of the origin of the malware and the discovery of the identity and motivation of its author hugely challenging, will inevitably mean that in the event of a serious and sophisticated cyber-attack, accusations of State sponsored involvement will persist. The uncertainties regarding attribution suggest that an arms control treaty is neither likely to increase the predictability in relations between potentially hostile States nor foster a more cordial atmosphere. If these are indeed the objectives sought by the proponents of an arms control treaty, there are perhaps more effective ways to secure such goals.[61]

In calling for the convening of the 1899 and 1907 Hague Peace Conferences, Russia was motivated by two factors: that, due to its economic weakness, it could not compete in the arms race with its rivals and in addition, its precious resources were being channelled into unproductive ends, namely, armaments.[62] This latter argument – that economic and social development should not be sacrificed for the benefit of military aggrandizement – was revived during the 1970s and 1980s under the UN rubric 'disarmament and development'.[63] In contrast with other weapons, cyber-weapons may paradoxically turn this argument on its head since the cost of enforcing a global prohibition may exceed any expected reduction in the level of risk. Moreover, as defensive tools acquire greater sophistication and capacity to detect and effectively respond to malicious codes, a complex regime to effectively monitor treaty compliance may prove far from cost-effective and even of subsidiary importance much in the same way that the utility of chemical weapons diminished considerably with the development of protective equipment.[64]

In the absence of agreement for an arms control treaty I consider in the following section whether existing LOAC rules offer an adequate basis for regulating cyber-weapons and their use. Since by contrast to other weapons, attacks using cyber-weapons are not primarily intended to produce a direct but rather an indirect kinetic outcome, does this require the re-evaluation of how LOAC rules pertaining to the means and methods of warfare apply? In particular I ask whether cyber-weapons are challenging to the law because they represent the essence of an ever

60 As experts have observed despite more than six years of negotiation on a proposed verification protocol, in 2001 the US withdrew its support although this came as little surprise given that the terms of the proposal were regarded by many of the participants as intrusive. See for example, Michael Moodie, 'Fighting the Proliferation of Biological Weapons: Beyond the BWC Protocol' 4 Disarmament Forum (2000) 33-42 and Kenneth Ward 'The BWC Protocol: Mandate for Failure' The Nonproliferation Review (Summer, 2004) 1-17. By contrast, verification under the CWC includes compulsory national declarations about relevant industrial and military activities, and a regime of routine inspections of declared industrial and military facilities. A particularly important feature is the provision for a 'challenge inspection' whereby a State party can request an inspection of any site in another State party at short notice; Robert Mathews and Timothy McCormack 'The influence of humanitarian principles in the negotiation of arms control treaties' IRRC No. 834, 30 June 1999.

61 '[U]ncertainty regarding attribution and the absence of common understanding regarding acceptable State behaviour may create the risk of instability and misperception'; Group of Governmental Experts report, 7 paragraph 7. 'The often low cost of developing malicious code and the high number and variety of actors in cyberspace make the discovery and tracking of malicious cyber tools difficult'; United States Department of Defense Cyberspace Policy Report: A Report to Congress Pursuant to the National Defense Authorization Act for Fiscal Year 2011, Section 934, November 2011, 8.

62 Goldblat, Arms Control: The New Guide to Negotiations and Agreements, 2.1.

63 Kalshoven, The Centennial of the First International Peace Conference: Reports and Conclusions, 97.

64 Mathews & McCormack, 'The influence of humanitarian principles in the negotiation of arms control treaties', 4.

inter-connected world in which conceptual and physical boundaries are being eroded, the very foundations upon which the law itself was constituted.

4. THE LAW OF ARMED CONFLICT: MEANS AND METHODS

It is self-evident that the ever-expanding use of information and communication technology in the critical infrastructure of States has created new vulnerabilities and opportunities for disruption, not least in the context of armed conflict.[65] Moreover, as the 2008 conflict in South Ossetia all too clearly demonstrates, adversaries will increasingly resort to strategies involving digital tools as an integral part of any military operation. Since the law of armed conflict applies to all situations that fulfil the criteria of an armed conflict, there is no coherent reason why the rules pertaining to the means and methods of warfare should not apply irrespective of methodology if the effects of deploying the malware produce the same outcomes as a kinetic weapon.[66] In fact, LOAC explicitly anticipates the emergence of new weapons and in doing so requires States to determine whether the use of any new weapon, means or method of warfare would be prohibited by international law.[67]

The law of weaponry which seeks to regulate both the means and methods of warfare can be traced back many centuries and its rules and principles are found in treaty and customary international law and in the growing body of case law generated by international courts and tribunals.[68] Although the St Petersburg Declaration is often cited for having been the first treaty to ban a particular type of weapon, a more important aspect of the Declaration is its preamble which, in setting out the reasoning behind the prohibition, articulates the general principles that have continued to inform the evolution of the law as it has confronted new means and methods of warfare.[69] The preamble reads:

> 'That the only legitimate object which State should endeavour to accomplish during war is to weaken the military forces of the enemy, […]
>
> That this object would be exceeded by the employment of arms which aggravate the sufferings of disabled men or render their death inevitable, [and]
>
> That the employment of such arms would, therefore, be contrary to the laws of humanity.'

65 GGE report paragraph 9, (A/65/201).

66 Commenting on the list of so-called weapons or 'fires', a senior military official indicated that the deployment of, for example, a computer virus would be governed by the same rules that apply to other military weapons, in other words, IHL; see The Washington Post, 1 June 2011.

67 Article 36 of Additional Protocol I provides, 'In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.'

68 As Greenwood notes, by the late 19th century, there was considerable support for the proposition that international law imposed some constraints upon the weaponry which a belligerent might employ; Greenwood, 'The Law of Weaponry at the Start of the New Millennium', 226.

69 The 1868 St Petersburg Declaration prohibited the 'use of explosive and incendiary projectiles weighing under 400 grammes which is either explosive or charged with fulminating or inflammable substances'. The convention did not prohibit the use of explosive projectiles per se as such weaponry was considered to be militarily necessary.

Two principles can be extrapolated from this. The first invokes the concept of military necessity, according to which only those weapons and means of combat which are necessary to attain the military purposes of war are permitted; this was subsequently given further weight with the incorporation of Article 22 of the 1907 Hague Regulations which explicitly provides that 'the right of belligerents to adopt means of injuring the enemy is not unlimited' (Article 22). More specifically, Article 23(e) prohibits the employment of 'arms, projectiles, or material calculated to cause unnecessary suffering'.[70] This principle is understood to prohibit both the use of weapons calculated to cause unnecessary suffering (means) and the use of otherwise lawful weapons if used in a manner that causes unnecessary suffering since to do so would serve no military purpose (methods).[71] A more recent expression of this principle is found in Article 35(2) of Additional Protocol I which provides that:

'It is prohibited to employ weapons, projectiles and materials and methods of warfare of a nature to cause superfluous injury or unnecessary suffering.'[72]

At first glance it is difficult to see how this prohibition would apply to cyber-weapons since the principle is concerned with the superfluous injuries and unnecessary suffering directly inflicted by the particular weapon on a combatant. Unlike in the case of other weapons, this principle would therefore appear to have little purchase on cyber-weapons insofar as their direct effects are concerned. If, however, the indirect effects of such weapons are taken into consideration it may be that malware designed to carry out specific tasks may potentially violate the principle, as for example where the destruction of medical data results in the provision of improper care of wounded combatants.[73]

Although the principle of unnecessary suffering has historically served as a basis upon which some weapons have been prohibited,[74] a compelling case can be made that the principle may make the use of cyber-weapons *more* likely. This is because if 'the essence of the unnecessary suffering principle is that it involves a comparison between different weapons in determining whether the injuries and suffering caused by a particular weapon are necessary', the cyber-weapon has the potential – to the frustration of those who wish to see its total prohibition – to 'outclass' all conventional weapons by inflicting least suffering.[75] Echoing the findings of a 1999 Department of Defense report, Denning observes 'instead of dropping bombs on an enemy's military communication systems, for example, cyber forces could take down the system with a computer network attack, causing no permanent damage and no risk of death or injury to soldiers or civilians. The operation would be more humane and should be preferred

70 The principle does not possess an absolute character because it only prohibits weapons that cause unnecessary suffering that cannot be justified by the military advantage that may be gained from its use.

71 'The only legitimate purpose of any use of weapons is the disabling of enemy combatants'; Dieter Fleck, Humanitarian Law in Armed Conflicts (1999) OUP, 121. The ICJ has described the principle of unnecessary suffering together with the principle of distinction as the two cardinal principles of international humanitarian law; Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996, paragraph 78.

72 See also rule 70, ICRC Customary International Humanitarian Law Study.

73 Arie Schaap, 'Cyber warfare operations: development and use under international law' 64 Air Force Law Review (2009) 121-174, 159.

74 For example, Declaration Concerning Expanding Bullets, 1899; Protocol on Non-detectable Fragments (CCW 1980, Protocol I); Protocol III (incendiary weapons primarily designed to set fire to materials, objects or to cause burn injury to persons; Gas Protocol 1925: prohibits use of chemical weapons directly against the enemy and to the toxic contamination of war-supply installations and food-stuffs; Biological Weapons Convention 1973; Chemical Weapons Convention 1993.

75 Greenwood, 'The Law of Weaponry at the Start of the New Millennium', 240.

over more destructive alternatives'.[76] Whether there would be an *obligation* on technologically advanced States to resort to digital options that cause less suffering if doing so does not reduce their military advantage remains far from clear.

The second general LOAC principle that unambiguously applies to cyber-weapons is the prohibition on the use of indiscriminate weapons or the indiscriminate use of any weapon. Once again, as with the principle of unnecessary suffering, these principles must be interpreted as applying to the intended indirect effect of the malware since 'the computer or network attacked is much less relevant than the systems controlled by the target computer or network […] [and] indeed the indirect effect is often the primary purpose of the attack'.[77]

As Greenwood notes, the principle of discrimination is a compound of three separate principles of customary international law: the principle of distinction, the principle of proportionality and the requirement to take all feasible precautions.[78] It therefore follows that if a particular cyber-weapon is incapable of being used in a way which enables a distinction to be drawn between military targets and civilians or civilian objects, it is inherently indiscriminate and therefore unlawful.[79] To the extent that a particular cyber-weapon can be deployed to attack a purely military objective and its destruction or neutralization provides a definite military advantage, the use of the malware would comply with the law.[80] Malware that cannot be contained or controlled and one that may cause injury to civilians or damage to civilian objects will constitute a prohibited indiscriminate weapon.[81] The proportionality principle which requires a balancing of the military advantages to be gained from an attack on a military target against the expected civilian harm and damage is even more difficult to evaluate for cyber-weapons given the interconnectedness of civilian and military networks.[82] This means that unless a rigorous assessment of the potential unintended consequences is conducted, a legitimate objective of attack may result in excessive collateral damage rendering the use of the malware unlawful in the circumstances.[83] As with the principle of unnecessary suffering, if the use of a cyber-

76 Denning, 'Obstacles and Options for Cyber Arms Controls', 7. See also Michael Schmitt, 'War, Technology, and International Humanitarian Law' HPCR Occasional Paper Series (2005), 55-56. See also DoD report As Assessment of International Legal Issues in Information Operations, May 1999, at 45: 'there is an obvious military interest in being able to interfere with an adversary's information systems, and in being able to protect one's own. Used as an instrument of military power, information operations capabilities have the significant advantage that they minimize both collateral damage and friendly losses of personnel and equipment. Their use may avoid unwanted escalation of a dispute or conflict'.

77 Technology, Policy, Law, 19.

78 Greenwood, 'The Law of Weaponry at the Start of the New Millennium', 242-243. See also ICRC Study, Rule 71.

79 'States must never make civilians the object of attack and must consequently never use weapons that are incapable of distinguishing between civilian and military targets'; Legality of the Threat or Use of Nuclear Weapons, paragraph 78.

80 Article 52(2) of Additional Protocol I requires that attacks are limited strictly to military objectives. It further provides that military objectives are limited to those objects which by their nature, location, purpose or use make and effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage.

81 Article 51(4)(c) of Additional Protocol I defines indiscriminate attacks as 'those which employ a method or means of combat the effects of which cannot be limited as required by this Protocol.' See also Schmitt, 'War, Technology, and International Humanitarian Law' footnote 114.

82 Technology, Policy, Law, 81 and 121-26.

83 '…cyberattacks aimed at military computer systems can have unforeseen consequences for civilian computers. Dams, nuclear power stations and civilian air traffic control all need computers in order to operate and to stay safe'; 'Cyber warfare and IHL' ICRC comment of 16 August 2011. For examples, see Schaap, 'Cyber warfare operations', 159.

weapon can secure the same objective as one involving the use of a kinetic weapon, it may well be the case that LOAC's obligation to take precautions in attack requires the State to the use of the former means because the risk of collateral damage and incidental injury would be considerably lower.[84]

Although the primary concern of this paper is with the law of weaponry and the means and methods of warfare, some comment is merited in respect of the rules on targeting. Although often conflated, because the principles that form the basis of a judgment as to whether a particular weapon or its use complies with the principle of discrimination are also relevant in respect of targeting, these two topics address separate questions. Be that as it may, discussions involving cyber-weapons consistently prompt two inter-related questions: the first concerns dual-use facilities; the second, whether the critical infrastructure of a State should be immune from a cyber-attack. The law regarding the former is fairly well settled since the question of targeting dual-use facilities is not unique to cyber-weapons.[85] Dual-use targets are understood as those that are used for both military and civilian purposes, as for example, power plants, oil and gas facilities, railroad and other transportation systems. In the digital age, this list has expanded to include, for example, computer networks of certain research facilities, air traffic control networks regulating both civilian and military aircraft, computerized civilian logistics systems upon which military supplies will be moved, electronic grid control networks, communications nodes and systems including satellite and other space-based systems.[86] For an object to qualify as a military objective, the target must 'make an effective contribution' to the enemy's military action; in other words, its destruction must provide a definite military advantage to the attacker.[87] The phrase 'make an effective contribution' is broad in scope and does not limit targets to only military objectives but to objects that make an effective contribution to the military; such objects may concurrently be of vital interest to the civilian population, as the examples above illustrate. However, before the target can be attacked, a proportionality test must be applied to ensure that the collateral damage to civilians or civilian objects is not excessive in relation to the concrete and direct military advantage anticipated.[88]

In a digitalized age, not only has there been an unprecedented increase in the number of potential targets that are dual-use in nature, but because networks are so interconnected, the resultant harm of an attack using malware is potentially enormous. Complying with existing LOAC rules that extend protected status to an area, or personnel or infrastructures will in practice be more difficult to observe since the interconnectedness of contemporary global society makes isolating specific interests that much more testing. For example, modern hospitals are 'highly networked facilities, dependent on telemedicine, and continuous retrieval of geographically remote information that is most likely stored in a data center that also houses other industrial

84 Article 57(2)(a)(ii) of Additional Protocol I requires those who plan or decide to pursue an attack to 'take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss or civilian life, injury to civilians and damage to civilian objects.'

85 For example, see Lawrence Greenberg, Seymour Goodman, Kevin Soo Hoo 'Information Warfare and International Law' National Defense University Press (1998) 12 and 37.

86 Schapp 'Cyber warfare operations' 156.

87 Article 52(2) of Additional Protocol I defines military objectives as 'limited to those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage'.

88 Article 51(5)(b) provides that an indiscriminate attack is one which 'may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated'.

and administrative data, and possibly even defense-related data.'[89] If such facilities are considered dual-use because they store defence-related data that is considered to make an effective contribution to the military action, would its destruction by the release of malware be lawful as long as the proportionality rule was satisfied? This has led some to suggest that certain critical infrastructures that are reliant on networks for their effective performance should be designated as immune from attacks by cyber-weapons.[90]

This cursory and partial[91] examination of existing LOAC rules suggests that the advent of the 'cyber-weapon' does not render the law obsolete. Nevertheless, as observed, 'applying pre-existing legal rules to a new technology raises the question of whether the rules are sufficiently clear in light of the technology's specific – and perhaps unprecedented – characteristics, as well as with regards to the foreseeable humanitarian impact it may have.'[92] Since malicious codes are designed to have different – and sometimes multiple – intended objectives, distinguishing between exploitation, intelligence-gathering, disruptions and conduct that is the prelude to something more serious will be challenging at best.[93] It may prove impossible to detect the existence of an armed conflict; destructive action including corruption, manipulation, or direct activity that threatens to destroy or degrade networks or connected systems may amount to use of force but it is unlikely to be regarded as an armed attack unless the outcome results in loss of life, injury and damage. But by contrast to an equivalent kinetic attack, the perpetrator may be difficult to identify since malware is often routed through servers in different countries. Perhaps more than any other domain of warfare, the unintended consequences in this new domain are the most troubling – both in respect of mistaken attribution and the level of harm that the deployment of a particular malware may inflict on the civilian population. While a cyber-weapons treaty or code of conduct will clearly not address all the most pressing issues pertaining to cyber warfare, it may assist in resolving some.

89 Karl Frederick Rauscher and Andrey Korotkov, 'Working toward rules for governing cyber conflict' EastWest Institute (2011) 22.

90 One problem that would first need to be overcome is that there is no consensus on what comprises the critical infrastructure. One definition is provided in the US Patriot Act, Section 1016(e), October 2001 which states: '[…] systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters'; cited by Rauscher & Korotkov, 'Working toward rules for governing cyber conflict', 12.

91 Issues that are not addressed in this paper include for example the protection of the environment, perfidy, neutrality.

92 'International Humanitarian law and New Weapon Technologies' Keynote address by Dr Jakob Kellenberger, ICRC, 34th Roundtable on Current Issues of International Humanitarian Law, San Remo, 8 September 2010. See also United States Department of Defense Cyberspace Policy Report: A Report to Congress Pursuant to the National Defense Authorization Act for Fiscal Year 2011, Section 934, November 2011, 8 which also concludes that 'the principled application of existing norms must be developed' to 'to clarify the application of norms and principles of customary international law to cyberspace'.

93 Technology, Policy, Law at 116. See also 2011 DOD strategy 9 on different categories of activities.

5. TENTATIVE CONCLUSIONS AND A POINT OF DEPARTURE

The prospect of a cyber arms control treaty cannot be dismissed outright since such treaties are always the product of a political choice governed by ever-changing priorities and perceived vulnerabilities. Once thought beyond effective proscription, mounting concerns over the horizontal proliferation of chemical weapons combined with the recognition by both the US and Russia that they did not need to retain their chemical weapons stockpiles following the Cold War culminated in the CWC.

That there is now greater willingness among States to enter into a dialogue on the implications of this new technology to armed conflict is to be welcomed. Where disagreement still persists is on the objectives that are being sought. For those that champion an arms control treaty, the critical question is whether there is any compelling reason why this particular weapon should be prohibited? As inferred above, the historical reasoning upon which other arms control treaties have been successfully negotiated and implemented have little bite. This is because in contrast to kinetic weapons, cyber-weapons are relatively inexpensive and widely accessible to non-state actors; the identity of the originating party behind a significant cyber-attack can be concealed with relative ease compared to that of a significant kinetic attack; it would be impossible to destroy all copies of the malicious code which may be stored in countless digital devices across the globe; and an effective inspection or verification mechanism is unlikely to materialise. Moreover, by contrast to other weapons that command public condemnation because they appear unambiguously indiscriminate or inflict unnecessary suffering, cyber-weapons are often regarded as a panacea that can achieve precisely the opposite: sanitize warfare and even prevent the use of kinetic force. Thus, rightly or wrongly, there is little public appetite to support a total ban.

In March 2009 Vladislav P. Sherstyuk, Deputy Secretary of the Russian Security Council, raised the possibility of a treaty to ban States from secretly embedding malicious codes or circuitry that could be later activated from a distance in the event of war. This comment raises an interesting question as to what constitutes a cyber-weapon, a question that I have attempted to answer above. However, if malware can be implanted that completely debilitates the armed forces of a State from resorting to kinetic force and does so without causing any casualties or damage, is such a device a weapon as understood under LOAC? The Stuxnet virus may have violated the principle of non-intervention and prohibition on the use of force but was its use governed by LOAC?

Such questions would probably not be answered by any multi-lateral agreement or code of conduct but a formal agreement of some form would provide a valuable framework within which to facilitate direct communication between States particularly in times of tension or crisis. The most serious threat that cyber space has engendered is the potential for armed conflict as a consequence of mistaken identity or alternatively, a misinterpretation as to intention.[94] The similarities between a cyber-attack and cyber exploitation mean that a targeted party may not

94 'After a call for a US-Russian bilateral high-level cyber security working group from Moscow in February 2011, US and Russian Delegations met in June with the goal of 'preventing misunderstanding and inadvertent escalation of cybersecurity incidents'; Joshua McGee, 'US-Russia Diplomacy – the 'Reset' of Relations in Cyberspace' Center for Strategic & International Studies, 5 August 2011 available at http://csis.org/blog/us-russia-diplomacy-reset-relations-cyberspace.

be able to distinguish between the two raising the risk of unwarranted or misinformed decisions in response.[95] This is compounded by the very nature of cyberspace such that there is now a need for more rapid responses creating higher levels of risk that a mistake will occur. A formal agreement may assist in addressing this problem possibly through a procedural mechanism or through the creation of an independent technical body that would assist with identifying sources of attack. A multi-lateral agreement would contribute to confidence-building, create an opportunity for States to affirm the applicability of LOAC principles and rules to cyber warfare and allow for the articulation of new norms should they be required. Such an agreement would also present an ideal opportunity to clarify the cyber lexicon and potentially allow for agreement to prohibit the use of cyber-weapons against critical infrastructures including for example, national power grids, financial markets or institutions, air traffic control systems.[96]

The modern law of armed conflict is founded on clearly delineated boundaries both conceptual and real. This vision of the world and the laws that were constituted upon this vision are now being challenged by cyber-space that thrives on the absence of boundaries. By their very properties, cyber-weapons are forcing us to re-evaluate our pre-conceptions about the nature of space, how we order our world, and the values which we most seek to preserve, not least in times of conflict.

95 Denning 2010; see also National Research Council, Letter report for the Committee on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy, 25 March 2010, 3. 'Cyberattack and cyber exploitation are technically very similar, in that both require a vulnerability, access to that vulnerability, and a payload to be executed. They are technically different only in the nature of the payload to be executed. These technical similarities often mean that a targeted party may not be able to distinguish easily between a cyber exploitation and a cyberattack.'

96 National Research Council, Letter report for the Committee on Deterring Cyberattacks, 21.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Internet as a CII - A Framework to Measure Awareness in the Cyber Sphere

Assaf Y. Keren
Communication and Cyber
Intelligence Solutions
Verint Systems
Herzliya, Israel
Assaf.Keren@verint.com

Keren Elazari
Communication and Cyber
Intelligence Solutions
Verint Systems
Herzliya, Israel
Keren.Elazari@verint.com

Abstract: With the increasingly vital role that the Internet plays in the personal lives of people across the world, Internet services have proliferated to such a degree that the Internet now equals countries' critical infrastructures in importance. In fact, some countries include Internet services in their legal framework for critical-infrastructure protection (CIP). In this paper, we take the view that the level of awareness and susceptibility to cyber attack can be measured by the level of maturity and development of the internet infrastructure of a country.
Using publicly available metrics, this study quantifies critical levels of Internet infrastructure across countries and proposes the cyber-attack susceptibility (CAS) index based on Internet usage, online services rendered, telecommunication infrastructure, and the human information-technology capital of each measured country. The information is used to further examine potential correlations between a country's critical Internet-infrastructure level and the country's ability to deal with cyber threats and the steps already taken by several high scoring countries in order to defend against attacks on Critical Infrastructure.

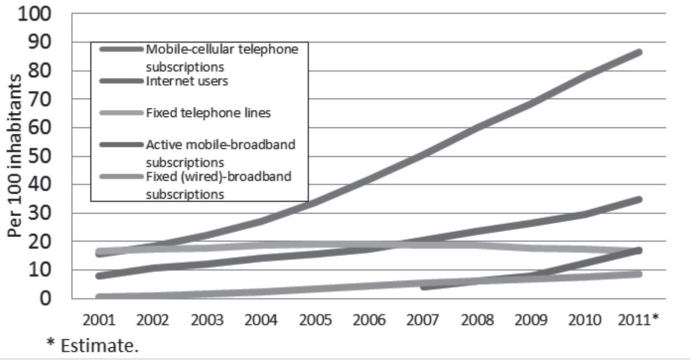
Keywords: *Internet; critical infrastructure; e-government*

1. INTRODUCTION

Over the last decade, the logarithmic scale of technology change has brought us into the information age in full force. One of the phenomena that this age has ushered in is the constant threat of cyber attacks - malicious attacks against computers and computer users. We are also witness to organized efforts by countries to develop the capability to attack other countries in the cyber sphere for the purpose of information gain or sabotage. This environment has led to the coining of the term *critical-information infrastructure*, an infrastructure that sustains life and must be defended against cyber attacks.

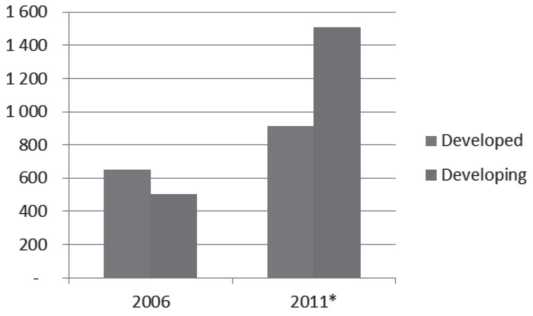
One of the biggest drivers of this was the creation and widespread proliferation of the Internet. Looking at data on global information and technology (ICT) developments [1], we can see constant growth in Internet and mobile-cellular telephone subscriptions over the 10-year period ending in 2011 (Figure 1).

FIGURE 1. GLOBAL ICT DEVELOPMENTS, 2001-2011.



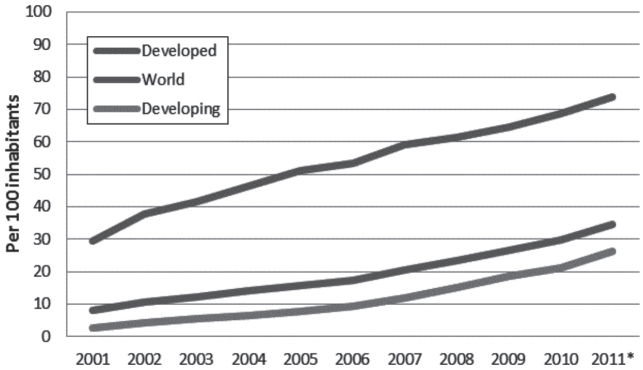
Moreover, the increase in Internet access is occurring not only in developed countries but also in developing countries [2] (Figure 2). As the latter race ahead with an ever greater number of Internet connections while disregarding issues such as proper infrastructure development and security, these countries' ability to protect critical infrastructure inside their borders is reason for concern.

FIGURE 2. INTERNET USERS BY LEVEL OF DEVELOPMENT[1], 2006-2011.



1 The classification of countries as developed or developing has been taken from the UN M49 standard.

FIGURE 3. INTERNET USERS PER 100 INHABITANTS, 2001-2011 [3]



But even more worrisome than the possibility of a cyber attack against critical infrastructures is that the Internet itself, we believe, has become a critical infrastructure of sorts. Somewhere in the rapid process of Internet development, some countries have become dependent on the Internet for providing a myriad of their services, such as e-government services, online banking, health services, life-saving instructions, and messages to the public.

This paper examines the Internet as a critical infrastructure, discusses levels of Internet connectivity and the provision of Internet-based services as meaningful indicators of the Internet as a critical infrastructure, and proposes a new framework for measuring countries' susceptibility to cyber attacks.

2. CYBER-ATTACK SUSCEPTIBILITY (CAS) INDEX

The cyber-attack susceptibility (CAS) index, as proposed by the authors of this paper, is composed of four indicators that help one gauge the level of Internet development in a country:

- The percentage of a country's population that uses the Internet. This indicator is based on figures from ITU (International Telecommunications Union) and other online sources [4].
- Online service index. The *United Nations E-Government Survey 2010* explains that "to arrive at a set of online service index values, the UN's research team assessed each country's national website as well as the websites of the ministries of education, labour, social services, health and finance….Among other things, the national sites were tested for a minimal level of Web content accessibility" [5].
- Telecommunication infrastructure index. This index is defined as "a composite of five indicators: number of personal computers per 100 persons, number of Internet users per 100 persons, number of telephone lines per 100 persons, number of mobile cellular subscriptions per 100 persons and number of fixed broadband subscribers per 100 persons" [5].
- Human capital index. The *United Nations E-Government Survey 2010* describes this index as "a composite of two indicators: adult literacy rate and the combined primary, secondary, and tertiary gross enrollment ratio" [5].

The online service index, telecommunication infrastructure index, and human capital index are used in the formula for the United Nations e-government development index (EGDI) [5]:

$$EGDI = (0.34 \bullet \text{online service index}) + (0.33 \bullet \text{telecom.infra.index}) + (0.33 \bullet \text{human capital index})$$

The CAS index score is the mean of the EGDI and the percentage of a country's population that is connected to the Internet:

$$CAS\ index = \frac{EGDI + \%\ of\ population\ connected\ to\ Internet}{2}$$

The four indicators (the percentage of a country's population that uses the Internet, the online service index, the telecommunication infrastructure index, and the human capital index) together give an idea of the degree to which a country's public participates in the Internet, the level of the country's governmental investment in Internet infrastructure and the technological literacy of the people, and the level of service that the country provides online. In other words, these indicators show the level of a country's Internet development and the reliance of its populace on Internet services.

The downside of a high level of connectivity and online services is that the latter are targets of cyber attacks. By this reasoning, countries that have a high CAS score are more susceptible to an attack that can leave the populace with a degraded Internet connection or none at all and that can result in a state of denial of service, the manipulation of content, or the theft of sensitive data.

3. CAS SCORES BY REGION

Table I lists the CAS scores by region and, in each region, the countries with the highest scores and the countries with the lowest scores.

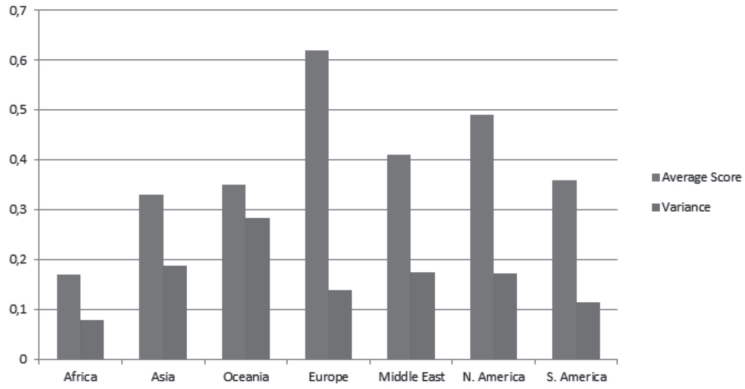
TABLE I: CAS SCORES BY REGION

Region	Mean CAS Score	Variance	Highest-Scoring Countries	Lowest-Scoring Countries
Africa	0.17	0.079	Tunisia, Mauritius	Chad, Niger
Asia	0.33	0.188	Rep. of Korea, Singapore	Nepal, Afghanistan
Oceania	0.35	0.283	Australia, New Zealand	Timor-Leste, Papua New Guinea
Europe	0.62	0.138	Norway, Netherlands	Albania, Bosnia and Herzegovina
Middle East	0.41	0.175	Israel, United Arab Emirates	Yemen, Iraq
N. America	0.49	0.171	United States, Canada	Haiti, Cuba
S. America	0.36	0.114	Argentina, Chile	Suriname, Nicaragua

Not surprisingly, as can be seen in Table I and Figure 4, the regions that have the highest

values for the CAS index are North America and Europe, the birthplace of the Internet. North Americans, especially in the United States and Canada, have become accustomed to the ongoing use of online information and have become reliant on the steady flow of information and services accessed via the Internet. North America's score is lower than Europe's only because of countries such as Cuba and Haiti, which are not as developed as the rest of North America; these countries also come into play in North America's higher variance of CAS scores.

FIGURE 4. CAS SCORES AND VARIANCE BY REGION



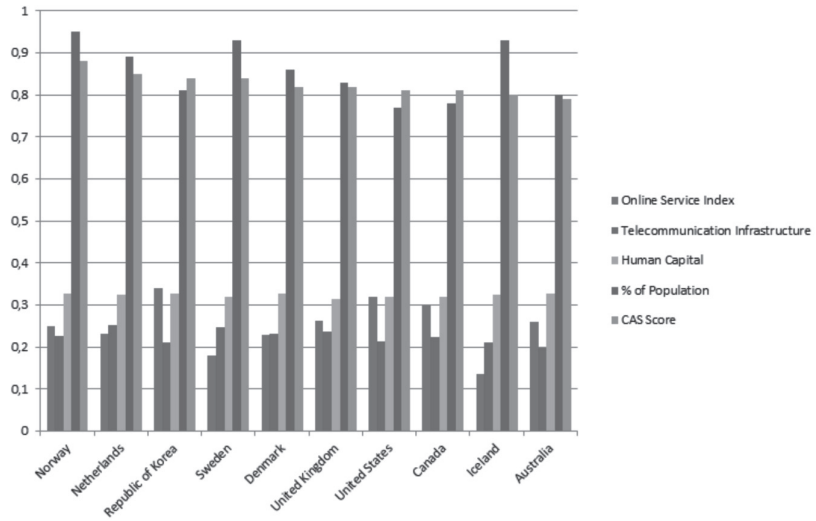
Again not unexpectedly, Africa boasts the lowest CAS score but also has a very low variance, indicating that the entire continent has relatively few online services and low rates of Internet connection and as such relies less on computers and Internet-based communication for everyday tasks and critical infrastructures.

Perhaps the most interesting region is Asia: one Asian country is among the 10 countries with the highest CAS index scores (Table II and Figure 5), but some of the countries that are the lowest CAS scorers are also in Asia (Table I). These two extremes result in a very high variance value (0.188) but can perhaps also affect the ability of Asian countries to cooperate as a region in the combating and mitigation of cyber threats.

TABLE II: TOP CAS-SCORING COUNTRIES AND THEIR COMPONENT INDEXES

Country	Online Services	Telecommunication Infrastructure	Human Capital	% of Population Connected to Internet	CAS Score
Norway	0.2504	0.2254	0.3262	0.95	0.88
Netherlands	0.231	0.253	0.3257	0.89	0.85
Republic of Korea	0.34	0.2109	0.3277	0.81	0.84
Sweden	0.1792	0.2482	0.32	0.93	0.84
Denmark	0.2288	0.2306	0.3278	0.86	0.82
United Kingdom	0.2634	0.2364	0.3149	0.83	0.82
United States	0.3184	0.2128	0.3198	0.77	0.81
Canada	0.3001	0.2244	0.3204	0.78	0.81
Iceland	0.1349	0.211	0.3238	0.93	0.80
Australia	0.2601	0.1983	0.3278	0.80	0.79

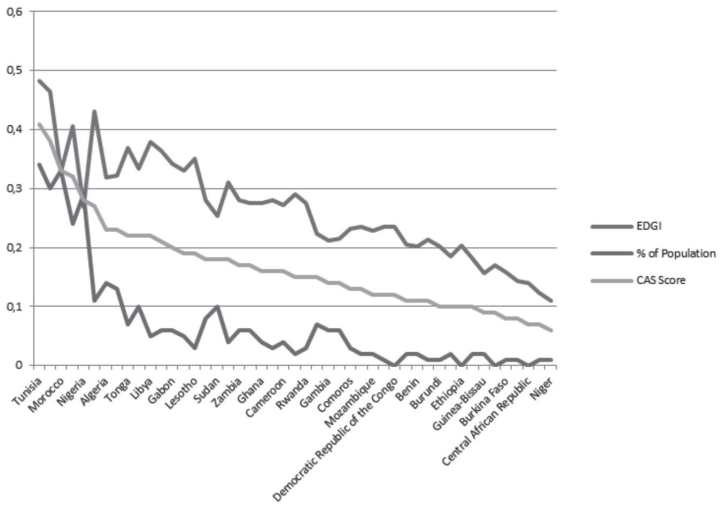
FIGURE 5. TOP CAS-SCORING COUNTRIES SHOWN WITH THEIR COMPONENT INDEXES



4. BREAKDOWN OF CAS SCORES WITHIN EACH REGION[2]

In Africa as a whole, the percentage of the population connected to the Internet is very low (Figure 6). Because the connection rate is below 10 percent for the majority of the countries in Africa, the EGDI is the most influential component of the CAS score for Africa. The CAS scores imply that African countries are not at high risk of cyber attacks.

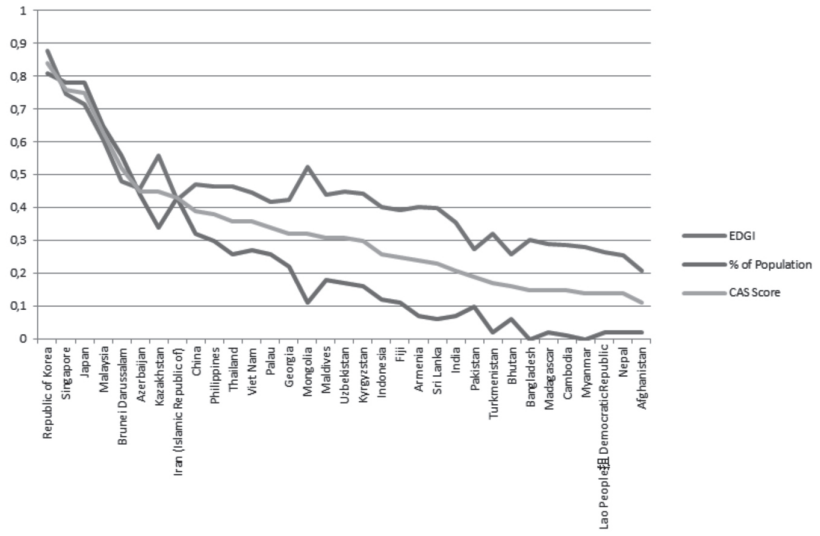
FIGURE 6. CAS SCORES IN AFRICA



2 For the actual CAS index scores of all the countries listed in this section, see the appendix. 6 Council of Europe. 2001. *Council of Europe - ETS No. 185 - Convention on Cybercrime*, [Online]. Available: http://conventions.coe.int/Treaty/en/Treaties/html/185.htm.

Asia is defined by a very large gap between the countries with the top four scores (the Republic of Korea, Singapore, Japan, and Malaysia) and the rest of the countries in the region (Figure 7).

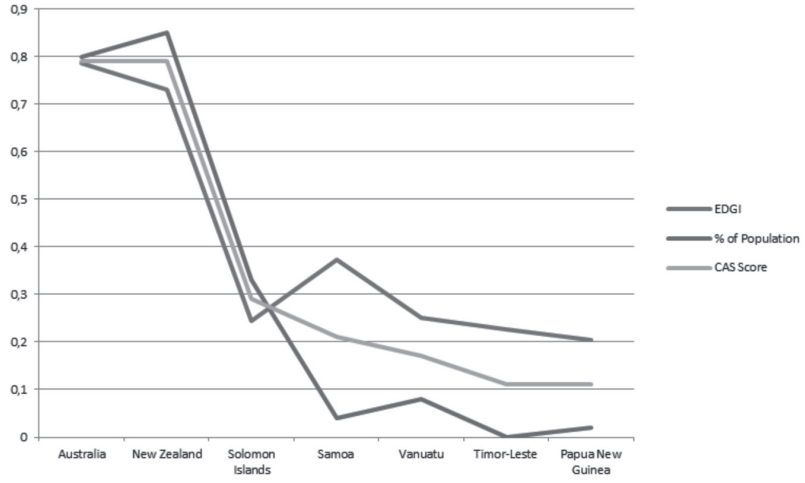
FIGURE 7. CAS SCORES IN ASIA



Cyber attacks are not only a greater threat in the top four countries of Asia, but these countries also have some of the most extensive programs in the world to deal with cyber threats, whereas the rest of the region is not as well positioned to handle such threats.

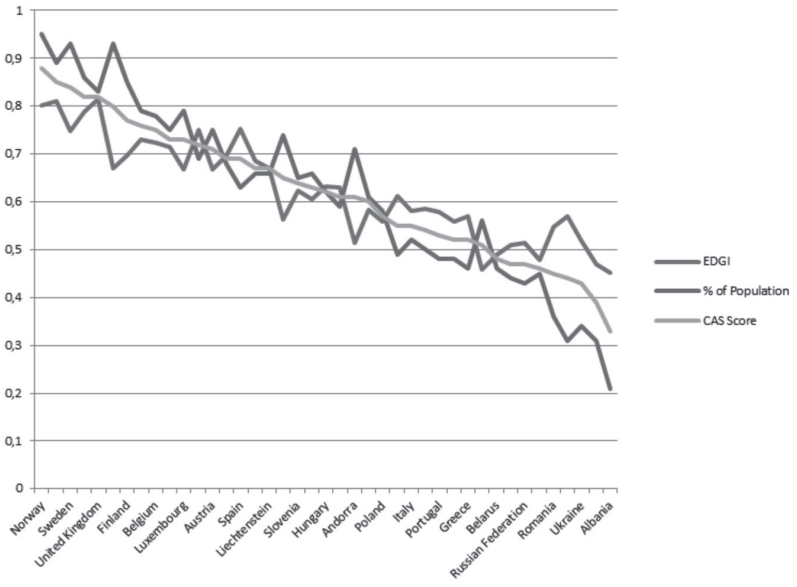
The situation in Oceania is similar to that in Asia: the top two countries (Australia and New Zealand) are much more advanced in Internet connectivity and online services than the rest of the region (Figure 8).

FIGURE 8. CAS SCORES IN OCEANIA



Europe has relatively high CAS scores across the region (Figure 9). There is a clear distinction between the northern part of Europe, which includes Scandinavia and western European countries such as Germany, the United Kingdom, and France and occupies the upper part of the CAS score table, and the southern and eastern parts of Europe, which occupy lower positions in the table.

FIGURE 9. CAS SCORES IN EUROPE

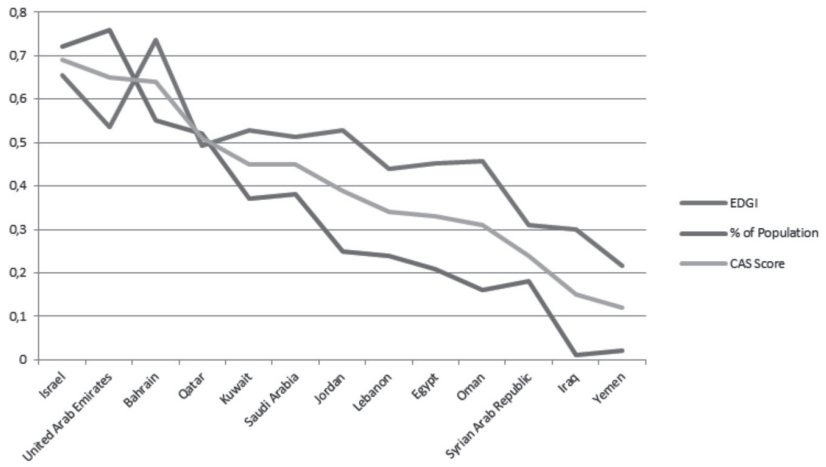


All in all, the comparatively even spread of the CAS scores in Europe enables European countries to more easily arrive at a better understanding regarding cyber threats and build regional structures, such as ENISA[3] (the European Network and Information Security Agency) and the Council of Europe Convention on Cybercrime [6], for facilitating cooperation and collaboration in the cyber sphere.

The scores of the countries in the Middle East demonstrate a clear division between the Persian Gulf States, Saudi Arabia, and Israel, on the one hand, and the rest of the region's countries, on the other (Figure 10).

3 http://www.enisa.europa.eu/

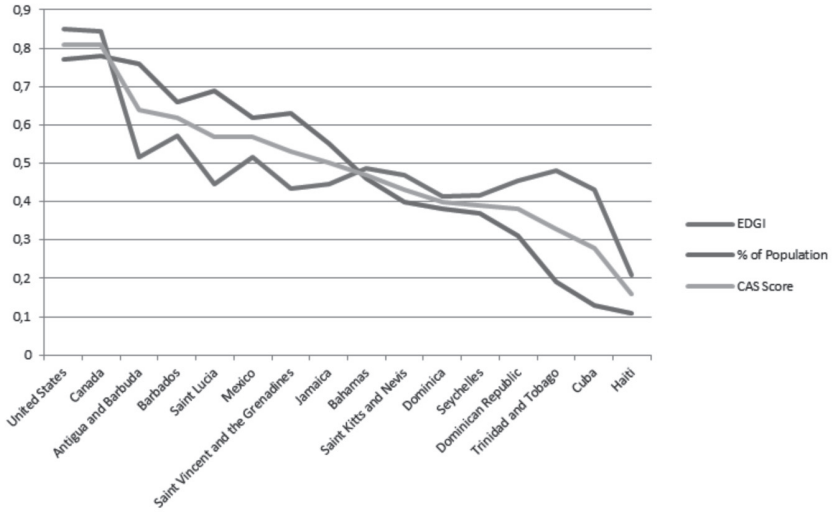
FIGURE 10. CAS SCORES IN THE MIDDLE EAST



Except for Israel, the United Arab Emirates, and Qatar, the CAS scores are influenced more by the EGDI scores than by the percentage of the population that is connected to the Internet, as clearly exemplified by Bahrain. These scores indicate that fewer people are connected to the Internet but the level of service that they obtain is quite good.

North America, like Asia, boasts a wide range of CAS scores (~0.2-~0.8), but unlike the graph for Asia, the North American curve slopes at a relatively steady rate (Figure 11).

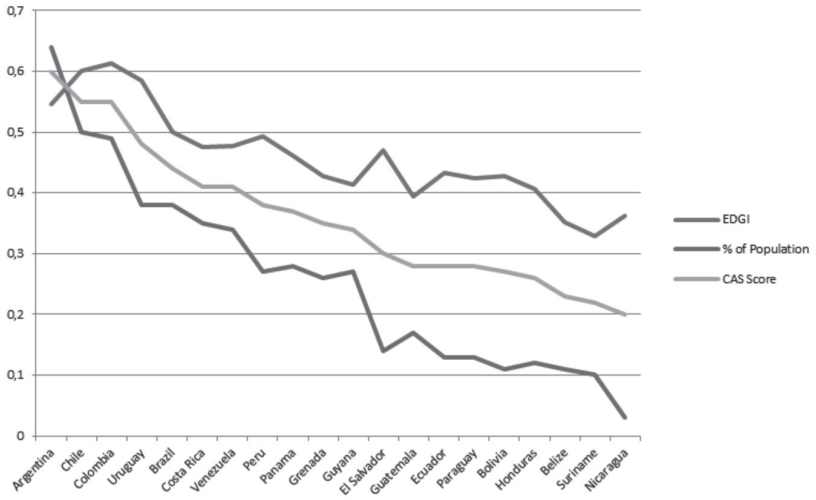
FIGURE 11. CAS SCORES IN NORTH AMERICA



Canada and the United States are the most dominant countries in the region and are also dominant globally. Both exhibit very high EGDI scores and a high percentage of the population connected to the Internet.

Other than Africa, South America is the least connected region in the world, with a very low percentage of the population connected to the Internet (Figure 12). However, the slope of the South American EGDI curve is unusually mild; there is only a small difference between the value for the country with the lowest EGDI score (Suriname) and the value for the country with the highest EGDI score (Colombia).

FIGURE 12. CAS SCORES IN SOUTH AMERICA



5. OVERVIEW OF CIP SCHEMES IN HIGH CAS-SCORING COUNTRIES

In this section, we review the critical-infrastructure protection (CIP) schemes of the seven countries with the highest CAS scores to demonstrate how countries that are at high risk of cyber attacks deal with such threats.

A. Norway

In Norway, the Ministry of Justice has overall responsibility for critical-infrastructure protection [7], with NorCERT as a supporting function for incident response. The Norwegian CIP commission's report identifies two types of systems to be protected. The first—critical infrastructure—includes "electrical power, electronic communication, water supply and sewage, transport, oil and gas, and satellite communication" [8]. The second covers "critical societal functions," which include "banking and finance, food supply, health services, social

services and social security benefit, the Police, emergency and rescue services, [and] crisis management" [8]. The report also indicates additional critical societal functions—"Parliament and government, the judiciary, Defence, Environmental surveillance and waste treatment" [8]—which were not examined by the commission.

B. The Netherlands

In 2005, the government of the Netherlands conducted a risk analysis that demonstrated the need to increase the protection of critical infrastructure [9]. As a result, the National Advisory Center on Critical Infrastructure (NAVI) was formed. With the establishment of the Center for the Protection of Critical National Information Infrastructure (CPNI.nl), NAVI's roles and responsibilities were transferred to that organization. Currently, critical infrastructure in the Netherlands is divided into twelve sectors: energy, telecommunications and ICT, drinking water, food, health, finance, surface water management, public order and safety, legal order, public administration, transport, and the chemical and nuclear industries [10].

C. Republic of Korea

Korea passed the Act on Information and Communications Infrastructure Protection in 2001 to establish a framework for the protection of highly sensitive networks in the country [11]. The supervision of critical-infrastructure protection is conducted by the Information and Communication Infrastructure Protection committee [12], which guides the various government ministries and agencies that handle the day-to-day protection of the critical infrastructure within their purview.

D. Sweden

According to a study by Germany's Federal Office for Information Security (BSI), Sweden's "critical infrastructure protection has been integrated into the general complex of national defense. Critical infrastructure protection is viewed as a combination of information assurance, critical infrastructure protection, defensive information operations and defensive information warfare" [13]. Instead of establishing one organization to be in charge of critical infrastructure protection, Sweden has divided the responsibilities among the Swedish Emergency Management Agency (SEMA), the Technical Competence Centre (TCC), and GovCERT.

E. Denmark

Denmark handles critical infrastructure through two bodies, the Danish Emergency Management Agency (DEMA) and GovCERT [14]. DEMA conducts risk analysis on an ongoing basis, and GovCERT, which belongs to the Ministry of Defence, is in charge of incident response assistance to selected critical-infrastructure owners [15].

F. United Kingdom

On February 1, 2007, the UK formed the Centre for the Protection of National Infrastructure (CPNI) from a merger of the National Infrastructure Security Co-ordination Centre (NISCC) and the National Security Advice Centre (NSAC). CPNI is responsible for providing "integrated security advice […] to organizations which make up the national infrastructure" [16]. CPNI is in charge of the protection of nine national infrastructure sectors: communications, emergency services, energy, finance, food, government, health, transport, and water.

G. United States

The United States has had a broad critical-infrastructure protection scheme in place since 1996. The protection plan was restructured under the Department of Homeland Security (DHS), as specified in the Homeland Security Presidential Directive no. 7 (HSDP-7) in 2003. Each of the protected sectors is under a sector-specific agency, and each agency has established a policy that addresses the various issues of the sector. The sectors that are under protection are agriculture and food; banking and finance; chemical; commercial facilities; communications; critical manufacturing; dams; the defense industrial base; education facilities; emergency services; energy; healthcare and public health; information technology; national monuments and icons; nuclear reactors, materials, and waste; transportation systems; and water.

6. THE CAS INDEX "TIPPING POINT" MODEL

It is our opinion that by using the CAS index, one can construct a model that indicates the "tipping point"—the point at which nations realize that they are susceptible to a critical threat of cyber attacks. A quick survey of the data presented earlier makes clear that no country with a CAS index below 0.21 has a scheme in place for critical-infrastructure protection. The country with the lowest CAS score that has a well-structured critical-infrastructure plan is India, which is anomalous because of the low percentage of its population, especially in rural areas, that is connected to the Internet (7 percent nationwide). On the other hand, all of the countries with a CAS score above 0.61 boast a working plan for critical-infrastructure protection, with Andorra as the first country without such a plan. Because of its size, Andorra might not need a comprehensive plan to combat cyber threats. The next country in the list without a critical-infrastructure protection plan is the Bahamas, with a score of 0.47, bringing us much closer to India's 0.21 score.

7. CONCLUSIONS

As the data show, the use of the Internet and technology is spreading across the globe. Developing countries are among the nations that are increasing their Internet connectivity at the fastest rates in the world. Along with technological advancement and Internet connectivity comes the threat of cyber attacks against critical infrastructure and hence the need for a framework to measure the susceptibility of countries to cyber attacks.

The CAS index framework can be used for numerous applications, from research to commercial to defense purposes, with more to come. In the future, we intend to expand our research to establish a solid mathematical "tipping point" model and hope that other researchers will use this framework for further investigation.

REFERENCES

[1] ITU. 2011. *Global ICT developments, 2001-2011* [Online]. Available: http://www.itu.int/ict/statistics.

[2] ITU. 2011. *Internet users, by level of development (2006-2011)* [Online]. Available: http://www.itu.int/ict/statistics.

[3] ITU. 2011. *Internet users per 100 inhabitants, 2001-2011* [Online]. Available: http://www.itu.int/ict/statistics.

[4] *World Internet Statistics* [Online]. Available: http://www.internetworldstats.com/stats.htm.

[5] UN Department of Economic and Social Affairs, *United Nations E-Government Survey 2010*, United Nations, New York, 2010.

[6] Council of Europe. 2001. *Council of Europe - ETS No. 185 - Convention on Cybercrime*, [Online]. Available: http://conventions.coe.int/Treaty/en/Treaties/html/185.htm.

[7] European Network and Information Security Agency, *Norway Country Report*, 2010.

[8] Norway Ministry of Justice, *Protection of critical infrastructures and critical societal functions in Norway*, 2006.

[9] European Network and Information Security Agency, *Netherlands Country Report*, 2010.

[10] Government of Netherlands. 2011. *Protecting Critical Infrasturcture*," [Online]. Available: http://www.government.nl/issues/crisis-national-security-and-terrorism/protecting-critical-infrastructure.

[11] Government of Korea. 2001. *Act on Promotion of Information and Communication Network Utilization and Information Protection* [Online]. Available: http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN025694.pdf.

[12] J. Jang, *The Current situation and Countermeasures to Cybercrime and Cyber-Terror in the Republic of Korea*.

[13] BSI. 2004. *Critical Infrastructure Protection: Survey of World Activity* [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Kritis/paper_studie_en_pdf.pdf?__blob=publicationFile.

[14] European Network and Information Security Agency, *Denmark Country Report*, 2011.

[15] Centre for the Protection of National Infrastructure. 2012. *The national infrastructure [Online]. Available:* http://www.cpni.gov.uk/about/cni

[16] GovCERT. 2011. *The following profile of the Danish GovCERT has been established in adherence to RFC-*2350 [Online]. Available: https://www.govcert.dk/gcdata/rfc2350_govcert.pdf.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

The Significance of Attribution to Cyberspace Coercion: A Political Perspective

Forrest Hare
Center for Peace and Security Studies
Georgetown University
Washington, D.C., United States
fbh5@georgetown.edu

Abstract: The question of cyber deterrence, or "What and how do you deter malicious actions in cyberspace?" has been hotly debated over the last few years. Stories of massive intellectual property theft and identity theft cases have surfaced in the Western news spurring several seminars and writings on the subject. Unfortunately, the discussion to date has not moved us effectively toward a comprehensive framework for building a coercion strategy. Most importantly, the debate has failed to accurately characterize the coercion challenge. In most cases confronting developed nations, the more pressing issue is not deterring an actor from choosing to conduct hostile intrusions in cyberspace but compelling the actor to stop conducting intrusions that already have been highly successful. Accurately recognizing the existing dynamic changes coercion calculations in several ways, such as the significance of positive attribution – an important component of coercion theory. Although the proposed coercion strategy framework in this paper will necessarily be less than comprehensive, one important outcome will be that the issue of unequivocal attribution may not be as critical as previously suggested.

Keywords: *cyber security, security studies, attribution, coercion, compellence*

"A difficulty with our being an unaggressive nation, one whose announced aim has usually been to contain rather than to roll back, is that we have not settled on any conventional terminology for the more active kind of threat."

THOMAS SCHELLING (1966)

The opinions expressed, or implied, in this article are solely those of the author, and are based on open sources. Under no circumstances may they be correlated to or perceived as the views of Georgetown University, the U.S. government in general, or the Department of Defense in particular.

1. INTRODUCTION

The question of cyber deterrence, or "What and how do you deter malicious actions in cyberspace?" has been hotly debated over the last few years. Stories of massive intellectual property theft and identity theft cases have surfaced in the Western news spurring several seminars and writings on the subject. Unfortunately, the discussion to date has not moved us effectively toward a comprehensive framework for building a coercion strategy. In fact, authors since the mid-1990s have been arguing that traditional deterrence theory is difficult to apply to current cyber threats (Alperovitch 2011; Harknett 1996; Libicki 2009). However, the debate continues because policy-makers remain unable to find efficacious answers to persistent, and immediate, threats in the domain. The aim of this paper is to advance the discussion forward by using a different perspective on coercion. Most previous writings have focused on the difficulties of applying traditional deterrence theory to the domain, such as the challenges to determining attribution. Most importantly, the debate has failed to accurately characterize the coercion challenge. In most cases of cyber conflict confronting developed nations today, the more pressing issue is not deterring an actor from choosing to conduct hostile intrusions in cyberspace but compelling them to stop conducting intrusions that already have been highly successful.

Accurately recognizing the existing dynamic changes coercion calculations in several ways. For example, it may alter the importance of positive attribution—an important component of coercion theory. To provide a different perspective on the significance of attribution, this paper proposes a cyberspace coercion framework that draws on insights from Schelling (1966), and modeling by Byman, Waxman, and Larson (1999) of RAND. The model by Byman et al. identifies a continuum of policy objectives, from deterring an actor from intruding in systems connected through cyberspace, to one of forcing an actor to stop threatening intrusions and remove malware implanted in critical infrastructure. Based on these objectives, the paper will highlight the relative importance of emplacing strong defenses, communicating retaliatory actions, achieving attribution, and executing effective responses to successful intrusions. It will build on the author's previous work (Hare 2010) regarding international cyber security dynamics to explore effective ways to "ratchet up the pain" necessary to compel actors to change their behavior in the domain. The goal of the paper is to revisit the issue of attribution through this framework and re-assess the importance of positive attribution. Though this proposed framework will necessarily be less than comprehensive, one important outcome will be to reveal that the issue of unequivocal attribution may not be as critical as previously suggested by many authors.

Before making the argument for a more appropriate coercion framework, I will establish a definition for the concept of national security in cyberspace that focuses the discussion on issues regarding international security relations. Thereafter I will provide a short review of the attribution problem in cyberspace as it is currently portrayed in the literature. This review will be followed by my argument on the need for a new coercion model for cyberspace. Using the new coercion model, I posit three potential coercive measures that will provide the opportunity to reassess the attribution problem. The paper concludes with points policy-makers should consider regarding attribution, given this new analysis framework.

2. THE CYBERSPACE THREAT TO NATIONAL SECURITY

In this section, I establish a definition for the concept of national security in cyberspace. This definition bounds the cyber security problem to security issues between nation-states. Using Buzan's (1991) concept of securitization and previous work by this author (Hare 2011), I first specify the public good of national security as, "that state in which the public of a nation is not threatened by something, or someone, that poses an existential threat."[1]

There are two primary ways this state of being can be threatened through cyberspace by adversarial nations and other malicious actors. First, a nation can suffer a threat from intrusions through cyberspace by either state or organized non-state actors against government, and select other, information systems to gain knowledge of national security value. Such activity, whether conducted by people intercepting bits and bytes of information or using their own eyes and ears, is generally considered espionage. Targets of such espionage could include the sensitive information systems of defense ministries or contractors that develop major weapons systems. Successful attacks would allow an adversary to counter a wide-array of national defense measures and they could justify governments using extraordinary measures to thwart such attacks, such as calls for increased deterrence options.

Second, a nation can suffer an existential threat from attacks and infiltrations through cyberspace by either state or organized non-state actors to degrade or disrupt critical infrastructure systems, both privately and publicly owned. For example, emplacement of malware and other disruptive software in the control systems used in the energy, transportation, or telecommunications sector could endanger many lives directly or thwart physical actions intended to defend national interests. Successful intrusions or attacks could also have a significant economic impact or cause a loss of life, and therefore again justify extraordinary counter-measures. Adding these two criteria to the definition of national security redefines the definition of the public good of national cyber security as the state of being in which the populace, governing institutions, and critical infrastructure are not threatened by:

- Attacks and intrusions through cyberspace, by either state or organized non-state actors, against government and select other information systems to gain knowledge of a national security value, or
- Attacks and intrusions through cyberspace, by either state or organized non-state actors, against critical infrastructure systems to degrade or disrupt such systems and cause a national security crisis.

This definition provides policy-makers with boundaries within which to develop a potential coercion strategy. An important component of this definition is the list of malicious actors against which a coercion strategy can be directed. Specifically, coercive actions would be taken to influence actors under direct control of the state or those acting with at least the tacit approval of the state. This second category could include paramilitary organizations, and contractors (Lewis 2011). In either case, state institutions of the adversary regime must be able to influence

1 Significant portions of this section are adapted from *The Interdependent Nature of National Cyber Security: Motivating Private Action for a Public Good* (Hare, 2011). For a more in depth discussion of the concepts in this section, please refer to this work.

both the action and inaction of the malicious actor. In other words, if coercion measures implemented by a threatened nation-state are to be successful, they must be directed at the authorities of the adversary nation who, in turn, must be able to exert their sovereign powers within their own territory. Actors such as patriotic hackers, criminals, and terrorists are much more difficult to coerce, as they are less susceptible to national power. However, since they seldom engage in the activities contained in the above definition of national cyber security, they are also a less significant threat to a nation's sovereignty (Lewis 2011).[2] With the problem thus bounded, I will now return to the calculus of coercion as it pertains to this domain.

3. THE CALCULUS OF CYBER COERCION

As mentioned in the introduction, several seminars and writings have been dedicated to a discussion of deterrence in cyberspace. Most have addressed the difficulties of applying deterrence theories in this domain, and many have focused on the challenges of achieving conclusive attribution of the malicious actors. For example, in his book *Cyberdeterrence and Cyberwar*, Libicki (2009) defines cyberdeterrence as an in-kind deterrence against attacks through cyberspace (p. 34). Using this definition, he highlights several issues that make such a strategy difficult to implement. His first question is, "Do we know who did it?" (p. 41). This is, of course, the issue of attribution. In his opinion, the victim must be able to convince third parties that the attribution is correct and, more importantly, the attacker must be convinced that the act will be correctly attributed to them. Libicki provides several examples of the difficulties in achieving conclusive attribution, based on the anonymity provided by the structure of the Internet and the indirect ways packets can be routed to their eventual targets.

Libicki (2009) even questions the idea that the beneficiary of an action would be its most likely instigator (i.e., *cui bono*), in that there are often many parties that could benefit from an attack. For example, several nations would be interested in intellectual property information from more advanced nations, and armed with this fact alone, it would be difficult to determine which nation had been responsible for a theft of intellectual property. Libicki also raises the possibility of false flagged operations being conducted to divert attention away from the malicious actor. Finally, he raises the practical concern that actions taken to demonstrate conclusive attribution to the international community or directly to an accused attacker will do nothing more than instruct the attacker on how to hide their activities more effectively. Clark and Landau (2010), in a paper specifically devoted to the challenge of attribution, state that "attribution is central to deterrence [...] [and] retaliation requires knowing with full certainty who the attackers are" (p. 25). With this imperative, the reader is left to assume that no deterrence strategy, whether intended to combat crime or defend a nation, can be effective without positive attribution.

Boebert (2010) breaks attribution down into technical and human components then discuss the barriers to achieving either forms, such as the proliferation of botnets and onion routing. He likens the problem of human attribution to that faced by any law enforcement agency that tries to solve a crime based on ballistic evidence. How do we prove who was at the keyboard at the time of an attack even if we identify the offending machine (Boebert 2010)? These authors and others have identified additional challenges to a successful deterrence strategy. Examples of other problems with deterrence in cyberspace include the difficulty of communicating a credible

2 A notable exception could be the events in Estonia in 2007 that will be discussed in this paper.

threat, the lack of clear red lines, and the risks of targeting innocent third parties with automatic responses (see Alperovitch 2011; Clark & Landau 2010; Harknett 1996; Libicki 2009; Lukasik 2010; Taipale 2010). With all of these challenges to a developing a robust deterrence policy, it is time to reassess the coercion problem in the cyber domain.

As the above review has shown, cyberspace presents many challenges when applying traditional deterrence theory. Nonetheless, the pressure to discuss deterrence in cyberspace has driven us to keep raising the attribution issue since it is perceived to be so critical to deterrence. To provide another perspective on attribution, I argue that we need to take different look at the problem of coercion: Is the problem really one of deterring an adversary from attacking us in cyberspace or is it a problem of compelling them to *stop* threatening intrusions that have thus far been very successful? There have been at least three instances of successful intrusion events that would support considering a different perspective.

The first such event was a broad intrusion set known as Ghostnet, which was first discovered by researchers in March 2008 and appeared to be continuing more than a year later (Deibert and Rohozinski, 2009). While the motivation and identity of the perpetrators has yet to be conclusively determined, the intrusion activities clearly targeted the communications systems of the office of the Dalai Lama, the Tibetan government-in-exile, and several non-governmental organizations affiliated with the Tibetan community (Deibert and Rohozinski 2009). Researchers with the Information Warfare Monitor identified an elaborate network of control servers and command servers that were being used to deliver and monitor targeted malware and exploit the information contained on over 1,000 computers in more than 100 countries. By the time the targeted communities discovered what was happening to them, there was no chance of deterring the perpetrator from conducting an act of cyber espionage. The problem instead became how to prevent the intrusion from continuing.

In 2011, a defense department official in the United States stated that, over the past few years, crucial files stolen from defense and industry data networks have included plans for missile tracking systems, satellite navigation devices, surveillance drones and top-of-the-line jet fighters (Shanker and Bumiller 2011). Once again, the military official was not aware of a potential threat to the critical data systems until the attack was well under way. The intrusions that the official is referring to may be continuing without their knowledge. At the very least, the intrusions clearly had occurred over an extended period without encountering any appreciable resistance.

Next, consider large-scale, Distributed Denial of Service (DDOS) attacks. In the case of some technologically advanced nations that have limited Internet-bandwidth, a DDOS against systems such as national banks and government communication systems could pose a significant risk to national security. For example, the small European country of Estonia experienced what it considered to be a debilitating series of DDOS attacks in 2007. These attacks occurred over several days, and there was little advance warning to give the nation's cyber defenders an idea of how broad or successful the attacks would be (Landler and Markoff 2007).[3] In this case, Estonia had no opportunity to develop, let alone communicate, a deterrent threat to any potential adversary. One would expect any potential victim to face this same challenge in deterring any

3 I would also argue that the sponsoring attackers probably did not know how broad or successful the attacks were while they were occurring.

potential DDOS attack that was not announced in advance. Given that a DDOS attack can be launched with virtually no warning and that doing so will greatly improve its effectiveness, such attacks can be expected to be used as first-strike weapons by any potential adversary.

Lastly, intrusions on the power grid in the United States have left behind software programs that could be used to disrupt the power production network, according to current and former national-security officials (Gorman 2009). According to a report in the *Wall Street Journal*, the intrusions were not detected immediately by the targeted power companies but by intelligence officials who identified pervasive espionage within the critical infrastructure sector (Gorman 2009). One would expect that several incidents similar to the four mentioned here have occurred but will remain unreported, due to their implications for national security in the targeted countries. A short survey of several government websites indicates that many countries are continually encountering intrusion activity on their government information networks (see, for example, Australian government n.d.). This success carries a message to hostile actors that such malicious activities will continue to be very rewarding, despite any strong rhetoric from victims. Based on the four incidents described here, I argue that we should revise the calculus of coercion. In so doing, we may find that attribution at the technical or legal level envisioned by previous authors may not be as critical as they conclude.

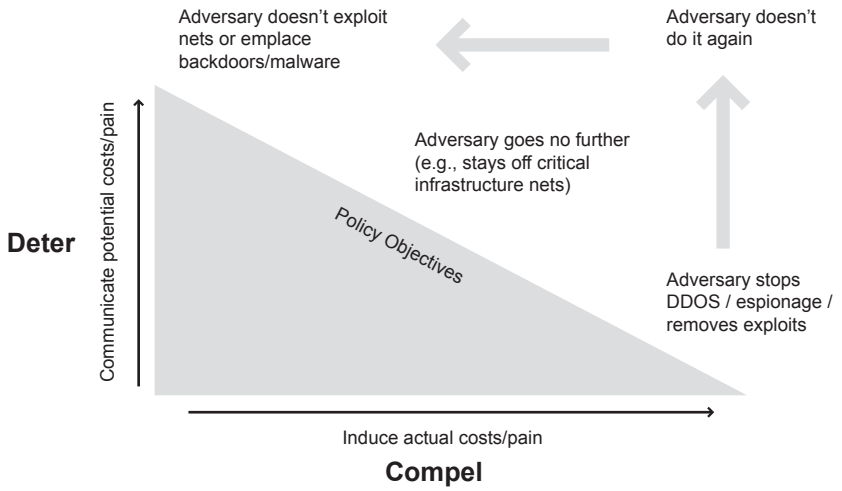
To develop a coercion model applicable to cyberspace based on the above evidence, we should take a fundamentally different approach than that taken by previous authors. If the malicious actors—adversary nation-states or their non-state proxies—have been conducting successful cyber espionage for an extended period of time, have sponsored no-notice DDOS attacks, or have already penetrated critical infrastructure control systems, then this author would argue that the coercion calculus is very different. The situation many developed countries now face is one that has been characterized by Nobel Laureate security strategist Thomas Schelling (1966) as one of *compellence*, not deterrence. According to Schelling, there are important distinctions between deterrence and compellence as components of a coercion strategy. The main differences are in the timing and the initiative. In a compellence situation, the attacker already has accomplished the offending action and the defender must take the initiative to respond, not just sit and wait. In other words, "The threat that compels rather than deters often requires that the punishment be administered *until* the other acts, rather than *if* he acts (Schelling 1966)." The compellence action taken must continue, or be believed to continue, until the offender responds favorably. There is no longer an ambiguous trip-wire that must be triggered before a threatened response is enacted. The line has been crossed, whether or not either party realizes it, and the offender has discovered the benefits have been worth the risk. The impetus is now on the victim to respond with a retaliatory action or assume the increased risk to national security. How much action is necessary will only be clear after the offensive behavior has been reversed or reduced to a level that is no longer considered threatening. However, in a deterrence situation, the defensive picture already has been painted. The adversary need not know the specific features of the painting, as long as no offensive act is committed. In fact, ambiguity may support deterrence. In a compellence situation, the picture must be painted for that specific situation and it must be clear to the offender what must be done, and by when, for the victim's coercive response actions to cease.

This simple change to the dynamic creates its own sets of challenges. First, to retain political

flexibility, it is no longer enough to leave the threshold or response ambiguous. The initiating party must communicate to the adversary that a retaliatory action is being taken in response to a specific action that is deemed hostile and is politically attributed to actors under the adversary's control. Whether or not the responsive action is to be initiated immediately, a deadline for compliance must be clearly articulated so the offender has no question as to when the offending act must cease. Second, not only must the terms be clearly communicated; the communication may have to be done in private so the adversary can avoid the appearance of having to back down. Inaction is easy to justify in a deterrence situation, as a would-be adversary can always claim other reasons for not conducting an action for which a victim threatens retaliation (Schelling, 1966). However, in a compellence situation, it is difficult for the offender to avoid the appearance of bending to the victim's will if the victim is to successfully influence the appropriate change in behavior. Third, in a compellence situation, the victim must develop a retaliatory action that will be effective and executable but that also can be stopped or reversed; otherwise there is no incentive for the offender to cease the offending behavior. Finally, on a positive note, if an effective compellence threat can be emplaced before the damage is too great, it may help overcome challenges faced in communicating the deterrence threat in the first place (Oh, you didn't know that was bad? Well, now you know, and if you stop, so will I). Given these challenges and opportunities, it is extremely important for the policy-maker to understand when they are presented with a compellence challenge.

Figure 1, which is adapted from a model produced by RAND researchers studying the use of air power as a coercive instrument (see Byman, Waxman and Larson 1999), depicts a framework for analyzing the coercion problem in cyberspace.

FIGURE 1. THE COERCION DYNAMIC AND POLICY



This figure shows the relationship between deterrence and compellence for achieving policy objectives to counter hostile cyber intrusions. The X and Y axes depict the relative weights of compellence and deterrence measures that must be taken to achieve exemplary policy objectives

presented along the slope of the triangle. For example, if the policy objective is to dissuade an adversary from executing threatening actions in cyberspace, then it may be sufficient to communicate to them the potential costs of taking such actions. As argued previously, several authors have assessed the many barriers to successfully communicating to the potential attacker that the costs of an offensive cyber action will outweigh the benefits. At the other end of the slope is a pure compellence situation in which the adversary has successfully penetrated and gained significant control over critical networks in a target country, or is currently conducting (or sponsoring) an effective DDOS attack against a victim's critical information systems. Up to this point, assuming the attack has been attempted, mere communication of the costs clearly has been unsuccessful. The only policy choice now available to the victim is to take the initiative to induce actual costs on the attacker in such a way that they respond to the counter measures and alter their offensive behavior.[4] If a victim finds itself in a situation where compellence measures have successfully removed an adversary from sensitive networks, it will most likely need to "keep the pressure on." A return to the *status quo* will be no more likely to deter the resumption of hostile actions by the adversary than it was before. Therefore, a policy mix must continue to induce some actual costs while also threatening the same or even stronger retaliatory measures, should the adversary return to exploit critical networks.

There are a few additional considerations I would like to address before proceeding with the analysis. First, the opaque nature of actions in cyberspace makes it difficult for the defender to know how far the attacker has penetrated and, therefore, exactly where they are on the policy slope.[5] Espionage will exist at some level and in all directions as long as the international system exists. If the victim finds itself in a situation where it sees that the attacker has penetrated to a certain point not viewed as an immediate threat to national security, then the appropriate coercion strategy may be a combination of deterrence and compellence measures. Second, when confronted with a compellence situation in cyberspace, the greatest policy challenge is to identify the appropriate costs or pain to be inflicted on the attacker to make them change their behavior in the desired manner (e.g., to get them off the critical networks). If the policy is restricted to taking retaliatory actions in cyberspace, then the victim's options may be limited. The counter to this point is that there is value in showing connectedness in response. A response that is closely connected in type and degree to the offensive action is easier to communicate to the offending party and to justify to an international audience (Schelling 1966). For example, launching cruise missiles in response to an act of cyber espionage may result in a proportional dollar loss to the offender, but it most likely will not be viewed by many as appropriate or sufficiently linked to the hostile cyber act that provoked the retaliatory measure. Finally, improper or poorly articulated goals can lead to a misapplication of pressure through coercive actions that will neither achieve desired results nor be measurable in any meaningful way. For example, it simply is not possible to stop all malicious actions in cyberspace, at least not with existing technology. However, it may be possible to influence the malicious behavior of nation-state actors to a measurable degree. Such a goal may be articulated as compelling the reduction of nation-state-sponsored espionage to a level that does not critically threaten national security.

Using the coercion model presented above, the policy-maker in the targeted country can more accurately identify where it is situated in the coercion dynamic. In this way, it can show whether

4 One characteristic of being at this point is that the victim had time to gather evidence of attribution from various sources. I will revisit this point later.

5 In the case of a DDOS attack, it may become clear very quickly, or it may not.

compellence measures are more appropriate than purely deterrent measures, such as establishing an already crossed red line. The next step will be to identify what actions in cyberspace the alleged offender will perceive as sufficiently threatening to favorably influence their behavior. Table 1, also taken from a previous work by the author (Hare 2010), helps to illustrate the types of malicious activities that various nation-states may consider most threatening to their national security.[6]

TABLE 1: CYBER THREATS AND TYPES OF NATION STATES

		Socio-political Cohesion	
		Weak	Strong
Power	Weak	De-stabilizing political actions in cyberspace, attacks on Internet infrastructure, criminal activities	DDOS and other major attacks on critical infrastructure [7]
	Strong	De-stabilizing political actions in cyberspace	Criminal activities in cyberspace

To get us back to the consideration of attribution, I will present two options that may be available to a country in the W-W, W-S, and S-S quadrants that is confronted with the challenge of compelling an actor associated with an S-W nation-state (bottom-left quadrant) to cease hostile cyber actions within the victim nation's sovereign territory. States in the bottom-left quadrant do not exhibit strong socio-political cohesion, and therefore perceive themselves to be vulnerable to threats to the idea of the state (legitimacy of the regime), its institution, and even its territorial integrity (Buzan 1991).[8] I will also discuss one option to be considered by a W-W, or W-S country (e.g., small European nations), when specifically confronted with a DDOS attack.

Option 1: Aggressive Engagement in International Forums

The first policy option could be a strong push via international forums for free and anonymous access to cyberspace by citizens of all nations. Actions to support this policy would entail government participation *en masse* at influential international conferences where regulatory, legal, and standards bodies debate potential measures to ease or restrict the freedom and flow of information in the domain. Adversarial countries would most likely be trying to persuade these same bodies to enact measures that would allow more state control on the flow of information. Participants from a targeted country would aggressively lobby other attendees and seek to dominate the agenda in a way that would send a clear message to the adversarial nation that actions taken at the conference are intended to counter them. This policy measure

6 This model was previously introduced for international cyber security discussions, but the perspectives are equally relevant for analyzing coercion strategies. Some of the cyber threats to national security suggested by this matrix, such as criminal actions to steal identity, may not be appropriate considerations for legitimate policy, but there are several options open to nation-states in each quadrant.

7 A distributed denial of service attack, or DDOS, occurs when many computers, usually surreptitiously controlled, are used to inundate a web server with requests and cause it to become overwhelmed to the point that service is denied.

8 For a complete discussion of this matrix and its contents see Hare, 2010.

has the advantage of being entirely executable within diplomatic channels, and being instantly "adjustable" meaning that the lobbying pressure can easily be reduced once the adversary's hostile actions have stopped. In addition to any overt actions taken, it would still be important for the victim nation to somehow communicate to the offender that these actions are in direct response to the hostile actions the victim has attributed to this adversary. For the actions to be effective as coercive measures, it must be made clear to the adversary nation that these diplomatic actions will stop once the threatening espionage and other hostile acts stop. If the adversary does not perceive that they have an opportunity to make the retaliatory actions cease, they may respond in unanticipated ways.

Option 2: Cyber Security for Dissident Organization in the Attacking Country

The second policy measure is significantly more aggressive. This option could be comprised of two related components enacted in steps. The first step would be to provide cyber security for a dissident organization countering the adversary regime. The security measures could entail providing hardware, software, and technical expertise to the dissident organization to protect their e-mail servers from the adversary's espionage and to protect the dissident organization's web presence from disruptions. The specific actions could be done in an overt manner to send a strong signal, or clandestinely to avoid causing an uncontrollable escalation in tensions. In either case, the adversary would have to be notified that the actions are being taken in response to perceived hostile acts they have sponsored. This policy action could be enhanced by communicating that if the adversary does not cease its hostile actions against the victim country's cyber assets, the victim will increase its coercive measures by conducting counter-espionage against the adversary and providing useful intelligence to the dissident organization. This second stage may be held in reserve to stress its compellence intent. However, its coercive effect can be highlighted by informing the adversary that some information has already been divulged to the dissident organization, such as information regarding the adversary's monitoring efforts of the dissident organization.[9] In any case, the adversary must be made to understand that the threat of increased counter-measures is not an empty one.

In both options, the actions would demonstrate clear connectedness and provide a potential deterrent to future cyber threats from the adversary. When a nation has demonstrated that it is willing and capable of taking action, it greatly increases the deterrent potential of the action. In all instances, it is important to signal to the target offender that these actions are taken in direct response to their hostile actions, and that the actions will cease once the offensive actions cease. The adversary may not respond at all if they don't realize that the victim nation has attributed the hostile activity to them (Libicki 2009). An unfortunate outcome of either set of measures would be that the adversary may respond to the actions in an escalatory manner. However, communicating the rationale for either option can be done in private, which would allow the offender to avoid the appearance of bending to the victim's will, an outcome that could be politically untenable for the adversary nation's regime.

Option 3: Hunker Down in the Face of a DDOS

Without any advance warning of an overwhelming DDOS attack, the victim nation will feel the effects almost immediately. There will be no period during which compellence actions like

9 This action must, of course, be balanced with the risk of exposing tradecraft.

those described above can be developed and deliberated. The unfortunate and probably the only choice will be to block originating addresses and endure the attack until the international spotlight can be turned on the likely perpetrator. According to Klimburg (2011), a senior advisor with the Austrian Institute of International Affairs, a small, Internet-dependent country can only hope to be successful with this tactic if it employs both a horizontal and vertical "whole of nation" approach to critical infrastructure protection. Such an approach would require a strong, public-private partnership that ensures a resilient and responsive infrastructure in the face of concerted attacks. Just as a country cannot predict when and where an earthquake will occur in sufficient time to evacuate all the buildings at the epicenter, a small, Internet-dependent country must make the "building strong enough" to resist attack. This option is less directly a compellence action than the first two options. In fact, the compellence theoretically does not come from resilience but from the international condemnation that would lower the attacker's international social capital to such a degree that they determine the cost to their international standing outweighs the benefits of a continued attack.[10] Therefore, the level of resilience necessary is related to the time it will take for the international community to observe and correctly characterize an event. In the case of Estonia in 2007, it took several days for them to receive support from other countries after the onset of anonymous attacks (Evron 2008).[11]

At this point, the reader may have expected to be presented with an option where a victim fights fire with fire, or a tit-for-tat response. However, several factors suggest that a DDOS counter-attack would be an ineffective response. First, it would most likely be enacted too late to be effective as a compellence measure because unless an action was preplanned, it would take time to determine which targets to strike and how to execute the attacks. Second, building extensive "botherds" of computers from which to launch an attack would be equally time-intensive, and relying on surrogate forces or criminal entities could be considered unethical options. Third, the attacking nation is most likely not as reliant as the victim on cyberspace for functions critical to national security and therefore would not be as heavily influenced by the attack. As identified in Table 1, the cyber actions most threatening to a nation in the S-W quadrant is the proliferation of information critical of the regime. As a result, a directly coercive action cannot be expected to influence the actions of the perpetrator of a DDOS attack enough to make them change their behavior. Regardless of the coercive measure taken, stronger defenses and increased resilience of the critical infrastructure must be a part of any strategy to increase the costs of conducting hostile actions in cyberspace and to help make retaliatory actions more effective.

4. ATTRIBUTION REVISITED

I now revisit the issue of attribution to determine its significance to the three policy scenarios presented above. As previous authors have done, I will address the technical and human components of attribution under each set of policy options.

10 I could write several paragraphs debating whether compellence or deterrence are even options to W-W and W-S type countries. However, the focus of this paper is on attribution, so I have entertained this scenario only to demonstrate that it conforms to the theories presented in the paper.

11 I acknowledge that it is impossible to determine if the international outrage ever did have an effect on the attackers since the attacks lasted for several more weeks. It is quite possible that they didn't cease until the attackers just got bored. This is one point that supports the previous footnote.

Attribution and Option 1: Engagement in International Forums

Aggressive lobbying for positions counter to the interests of certain regimes is not considered a hostile act under existing international conventions. Therefore, this coercive measure is not difficult to justify to internal or international audiences. There is no expectation that the victim country must explain to third parties that its action is a response to offensive actions by the adversary. Therefore, there is little need to attribute the initiating action to an international audience. However, a certain degree of human attribution is still necessary to make this option effective. The victim country must at least discover some correlation to the regime of the targeted nation-state and offer evidence that the offending actor is an entity over which the regime can exert some control if it so chooses. If the actor is one over whom a national regime cannot exert its sovereign power, then there is little chance the measure will achieve the intended coercive effect. Some evidence of this correlation may need to be conveyed to the adversary regime (potentially confidentially) to help the offender understand the link between its actions and the victim's response. Some amount of attribution may also be important when the adversary regime feigns ignorance of the event or if the adversary claims in an international forum that the aggressive diplomatic activity is an unprovoked assault. Technical attribution, even though it may have been achieved to support a determination of human attribution, is much less important in this situation. There is no policy requirement to tie an offending action to a specific machine, as the response measure is not tied to any specific machine. Ultimately, the burden is on the regime of the adversary nation to bring the hostile behavior to a halt. To do so, the regime will need to direct specific actors to alter their behavior. As long as the regime is aware of the responsible parties, knowledge of the specific networked entities used to conduct the hostile acts is of less importance.

Attribution and Option 2: Cyber Security for a Dissident Organization

This measure will clearly be considered threatening by the regime of an S-W nation. It also could be perceived by third parties as a direct challenge to the sovereignty of the targeted nation and therefore create concern in the international community. This concern will be greatest if any actions associated with the policy measures involve conducting a cyber operation within the territory of the targeted nation. In this case, there is a real possibility that the adversary nation will try to paint itself as the offended party and complain to the international community that it is experiencing an unprovoked attack on its sovereignty. Moreover, the victim nation initiating the response action must strive to demonstrate to the adversary regime that the measures are intended to be directly connected to and in response to actions determined to have been taken by the adversary.

Because of these two concerns, there is a greater requirement for some level of attribution that can be demonstrated to the adversary and, if events become public, to the international community. As with the first policy option, the focus of attribution must be the human or organization that perpetrated the offensive intrusions. Any evidence with which the victim chooses to demonstrate this linkage can be used to support attribution. It can consist of technical data, or intelligence gathered by multiple means. The argument to support attribution can also consider other events that may show a *cui bono* reason why the adversary was the most likely perpetrator of the hostile act. The duration of events may help determine attribution. An intrusion event of enduring nature may have increased the threat to national security, but it also gives the victim

with more opportunities to gather evidence of attribution using various means. Compiling several sources may allow the victim to obfuscate the specific details of any one source and avoid the unintended consequence of providing positive feedback to an adversary on how they can improve their intrusion tactics (Libicki, 2009). One of the sources could be technical attribution. Since the threatening espionage will have required two-way communications over an extended period to retrieve intelligence, it may provide more opportunities for a victim to identify valid source codes and overcome one of the largest barriers to attribution identified by Clark and Landau (2010)—the multi-stage attack. The espionage-motivated intrusions will most likely be against several information system targets. If the victim unravels the intrusion events, the complexity may provide an aggregation of evidence from several systems (Lewis 2011). Although the challenges of attribution do not change in this situation, the opportunities to achieve it may increase.

Attribution and Option 3: Hunker Down in the Face of a DDOS

As in option 1, enacting a strong defense that ensures the availability of services in the face of a DDOS attack and the regeneration of data in the case of a server crash would not be considered hostile acts. These are purely actions of critical infrastructure protection. As such, the only attribution necessary is that required to block offending intellectual property addresses at the target. In most cases, this technical attribution back to the last hop is easy to achieve. Unlike a case of espionage or more surgical cyber exploitation of critical infrastructure, there would be less exposure of state secrets if the event were publicized. In that case, it would be less politically risky to invite the broader community to advise in the defense of the nation. Therefore, cyber security experts from around the world could be invited to participate in the defensive actions and to help build a clearer picture of attribution. The necessary level of attribution is contingent on the case to be made to the international community. As in the previous case, the argument to support attribution can also consider other events that may show a *cui bono* reason why the adversary was the most likely perpetrator of the hostile act. For example, political disputes between Estonia and a neighbor nation provided the Estonian responders with a clear suspect. Also, the duration and level of a DDOS attack may help the defenders compile a sufficient level of human attribution based on the cumulative technical attribution. In the attacks on Estonia, on-line forums in certain communities were abuzz with discussions and instructions on how to participate in the attacks, which contributed to the determination of human attribution (Herzog 2011). Whether or not a particular regime is directly engaged in an activity, the international community may still consider it culpable and responsible, either legally or politically, for influencing the malicious behavior of actors under its influence. If a regime is unwilling to deal with the malicious actors or claims it is unable to do so, it could cause that regime to lose political capital. However, the policy-maker in the victim country must be willing to accept the fact that the issue of attribution is irrelevant if a loss of stature in the international community is irrelevant to the probable perpetrator.

5. CONCLUSIONS

There is no question that the anonymity and ease of international interaction in cyberspace increases opportunities for malicious activity. However, the coercion challenge is no more difficult in cyberspace than in other domains. The amount of evidence required to support an attribution argument will depend on the political situation at the time of the response action and the adversary's receptiveness to the victim's efforts to link the two actions. If there are many other factors pointing to a specific adversary as the likely instigator (the *cui bono* test), then that adversary will most likely be the focus of attribution and the international community will more readily support a claim of attribution without specific evidence.

In this paper, I have argued that unequivocal attribution is not required to enact a retaliatory measure and that attribution may be determined only after the measure is enacted successfully. However, for the compellence measure to be successful, the adversary must know that the victim has attributed the hostile actions to them, that the compellence measure is in retaliation for the offending action, and that the pain will stop only after the adversary has complied with the victim's demands. Confident assessment of human attribution will strengthen the effect of coercive responses. While the anonymity provided by cyberspace allowed the offender to conduct a threatening act that is not visible to others, it also enables a flexible coercion strategy. For example, it allows the communication and application of the compellence measure to be conducted privately. The benefit here is that the victim can plan its response actions with less concern about the influence of third parties or the demands of conclusive attribution.

For the three potential policy options discussed in this paper, attribution is a useful but not a required component of a coercion strategy. At the international level, national decisions are based on political considerations over legal ones. In a legal situation, attribution may be a requirement, but in a case of political calculus, attribution is one factor that must be balanced with all other political considerations of national security.

ACKNOWLEDGEMENTS

I would like to thank Col. David Fahrenkrug for giving me the idea of changing the coercion discussion from deterrence to compellence (and lending me the books to learn what that means). I would also like to thank my colleague Jeff Goldman for insightful comments on the first draft of my paper and the assigned reviewers for having forced a more international perspective in my presentation.

REFERENCES

Alperovitch, D. 2011. "Towards Establishment of Cyberspace Deterrence Strategy." In 3rd International Conference on Cyber Conflict (ICCC), 1–8. Institute of Electrical and Electronics Engineers.

Australian Government. "Mitigating the Cyber Threat". Government Document. Connecting with Confidence: Optimising Australia's Digital Future. http://cyberwhitepaper.dpmc.gov.au/white-paper/security-and-resilience-in-the-online-environment/mitigating_the_cyber_threat.

Boebert, W. Earl. 2010. “A Survey of Challenges in Attribution.” In Proceedings of a Workshop on Deterring Cyberattacks:Informing Strategies and Developing Options for U.S. Policy, 41–52. Washington, D.C.: The National Academies Press.

Buzan, Barry. 1991. People, States, and Fear: The National Security Problem in International Relations. 2nd ed. Boulder: Lynne Rienner.

Byman, Daniel, Matthew C. Waxman, and Eric Victor Larson. 1999. Air Power as a Coercive Instrument. Santa Monica, Ca: Rand Corporation.

Clark, David, and Susan Landau. 2010. “Untangling Attribution.” In Proceedings of a Workshop on Deterring Cyberattacks:Informing Strategies and Developing Options for U.S. Policy, 25–40. Washington, D.C.: The National Academies Press.

Deibert, Ronald, and Rafal Rohozinski. 2009. Tracking Ghostnet. Intrusion Analysis. Information Warfare Monitor. Toronto: Centre for International Studies, University of Toronto. http://www.scribd.com/doc/13731776/Tracking-GhostNet-Investigating-a-Cyber-Espionage-Network.

Evron, Gadi. 2008. “Battling Botnets and Online Mobs.” Georgetown Journal of International Affairs 9: 121.

Gorman, Siobhan. 2009. “Electricity Grid in U.S. Penetrated By Spies.” Wall Street Journal, April 8, on-line edition, sec. Technology.

Hare, Forrest. 2010. “The Cyber Threat to National Security: Why Can’t We Agree?” In Conference on Cyber Conflict Proceedings 2010, 211–226. Tallinn, Estonia: CCD COE Publications.

Hare, Forrest. 2011. “The Interdependent Nature of National Cyber Security: Motivating Private Action for a Public Good”. Doctoral Dissertation, Fairfax, Va: George Mason. http://u2.gmu.edu:8080/handle/1920/6312.

Harknett, Richard. 1996. “Information Warfare and Deterrence.” Parameters (Autumn 2006): 93–107.

Herzog, Stephen. 2011. “Revisiting the Estonian Cyber Attacks: Digital Threats and Multinational Responses.” Journal of Strategic Security 4 (2) (July 1). doi:10.5038/1944-0472.4.2.3. http://scholarcommons.usf.edu/jss/vol4/iss2/4.

Klimburg, Alexander. 2011. “Cyber Security Und Schutz Kritischer Infrastrukturen”. Newsletter. Newsletter der GIT Gesellschaft für Informations- und Kommunikationstechnik im OVE. http://git.ove.at/newsletter/GIT_Newsletter_05_2011.htm#klimburg.

Landler, Mark, and John Markoff. 2007. “Digital Fears Emerge After Data Siege in Estonia.” The New York Times, May 29, sec. Technology. http://www.nytimes.com/2007/05/29/technology/29estonia.html.

Lewis, James. 2011. “Rethinking Cybersecurity – A Comprehensive Approach” presented at the Sasakawa Peace Foundation, September 12, Tokyo. http://csis.org/publication/rethinking-cybersecurity-comprehensive-approach?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+CSIS-Cybersecurity-Related-Publication+%28Cybersecurity+-+Related+Publication%29.

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

Lukasik, Stephen. 2010. “A Framework for Thinking About Cyber Conflict and Cyber Deterrence with Possible Declaratory Policies for These Domains.” In Proceedings of a Workshop on Deterring Cyberattacks:Informing Strategies and Developing Options for U.S. Policy, 99–121. Washington, D.C.: The National Academies Press.

Schelling, Thomas C. 1966. Arms and Influence. Yale University Press.

Shanker, Thom, and Elisabeth Bumiller. 2011. “Hackers Gained Access to Important Files, Pentagon Says.” The New York Times, July 14, sec. World. http://www.nytimes.com/2011/07/15/world/15cyber.html.

Taipale, K. A. 2010. “Cyber-Deterrence.” SSRN eLibrary (April). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1336045.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

The Militarisation of Cyberspace: Why Less May Be Better

Myriam Dunn Cavelty
Center for Security Studies
ETH Zurich / Swiss Federal
Institute of Technology
CH- 8092 Zürich, Switzerland
dunn@sipo.gess.ethz.ch

Abstract: Cyber security is seen as one of the most pressing national security issues of our time. Due to sophisticated and highly publicised cyber attacks, it is increasingly framed as a strategic-military concern and many states have or at least want to acquire offensive cyber "weapons". The aim of this paper is to show that particular ways of framing threats are not only a matter of choice but also come with political and social effects. Focusing on the strategic-military aspects of cyber security means subjecting it to the rules of an antagonistic zero-sum game, in which one party's gain is another party's loss. This invokes enemy images even though there is no identifiable enemy, centres too strongly on national security measures instead of economic and business solutions, and wrongly suggests that states can establish control over cyberspace. This creates an unnecessary atmosphere of insecurity and tension in the international system - one that is based on misperceptions of the nature and level of cyber risk and on the feasibility of different protection measures in a world characterised by complex, interdependent risk. While it is undisputed that the cyber dimension will play a substantial role in future conflicts of all grades and shades, threat-representations must remain well informed and well balanced at all times in order to rule out policy (over-) reactions with unnecessary costs and uncertain benefits.

Keywords: *cyber security, cyber war, vulnerability-based planning, threat framing*

1. INTRODUCTION

As a result of increasingly sophisticated cyber incidents and intensifying media attention over the last few years, cyber security issues have moved in two directions: upwards, from the expert level to executive decision-makers and politicians; and horizontally, advancing from mainly being an issue of relevance to the US to one that is at the top of the threat list of more and more countries. On the national level, several governments have released or updated cyber security strategies in 2011.[1] Internationally, there is heightened attention on the strategic-military aspects of the problem – indicated by the growing number of conferences that address the issue,

1 Examples are France, Germany, India, the Netherlands, the United Kingdom, the United States and Switzerland.

efforts to obtain offensive capabilities, and attempts to come to an international agreement on the military (mis)use of cyberspace.

Though the heightened attention on cyber threats coupled with the overall sense of urgency to find viable political solutions could easily create the impression that policy-makers are confronted with an altogether 'new' issue, the current episode is just the latest development in the three to four decade long history of cyber threats. From the very beginning of the cyber threat story in the 1980s, there was a national security connotation to it (Dunn Cavelty 2008). However, that particular focus has intensified over the years, in parallel to society's increasing 'cyberification' and the overall impression that cyber incidents are becoming more frequent, more organised, more costly, and altogether more dangerous.

The establishment of cyber threats as a focal point of the current national security debate amongst Western states can be seen as a confluence of two interlinked and mutually reinforcing factors: the perception that modern societies are exposed to an ever-increasing number of potentially catastrophic vulnerabilities (Furedi 2008), and the perception of an increasing willingness of dangerous actors to ruthlessly exploit these vulnerabilities. This pervasive sense of vulnerability comes with a heightened sense of dread and urgency; and has led to a propensity to 'militarise' the cyber security debate.[2] The (unintended side) effects of this particular threat framing are the focus of this paper.

The aim is to show that particular ways of framing threats or risks are not only a matter of choice (within certain boundaries) but also come with political and social effects. Zooming in on the strategic-military aspects of cyber security means subjecting it to the rules of an antagonistic zero-sum game, in which one party's gain is another party's loss. This invokes images of a supposed adversary even though there is no identifiable enemy, is too strongly focused on national security measures instead of economic and business solutions, and wrongly suggests that states can establish control over cyberspace. In all, this creates an unnecessary atmosphere of insecurity and tension in the international system, which is based on misperceptions of the nature and level of cyber risk and on the feasibility of different protection measures in a world characterised by complex, interdependent risk.

To make this argument, the paper first describes three alternative ways of framing cyber security. This includes looking back to the 1990s when a well-balanced set of policy-responses took shape that were characterised mainly by a focus on the protection of critical infrastructures by technical means and a limited role of the military. The second subchapter examines recent developments and occurrences (spearheaded by Stuxnet, the Industry-sabotaging super-worm) that have given rise to an increasing focus on and attempts to acquire offensive cyber means. The third section critically assesses both the underlying assumptions behind this trend and the detrimental effects it has on the overall level of security. It is suggested that moving away from the propensity to think about worst-case scenarios and focusing on everyday occurrences like cyber crime and cyber espionage is the solution. The chapter concludes by arguing that military countermeasures will not be able to play a significant role in cyber security due to the nature of the information environment as well as the nature of the threat.

2 I use the term militarisation loosely, to connote the particular focus on the strategic-military dimen-sions of a problem and the adoption of something for use by or in the military.

2. ALTERNATIVE CYBER-IN-SECURITY FRAMINGS

In the evolution of the cyber security debate, we can distinguish between three different, yet closely interrelated and reinforcing discourses.

TABLE 1: THREE ALTERNATIVE CYBER DISCOURSES

	Technical	Crime-Espionage	Military / Civil defence
Main actors	• Computer experts • Anti-virus industry	• Law enforcement • Intelligence community	• National security experts • Military • Civil defence establishment / Homeland security
Main referent object	• Computers • Computer networks	• Private sector (business networks) • Classified information (government networks)	• Networked armed forces (military networks) • Critical (information) infrastructures
Main Threat	• Malware • Network disruptions • Hackers (all kinds)	• Advanced Persistent Threats • Cyber Criminals • Cyber mercenaries • States (foreign intelligence)	• Catastrophic attacks on critical infrastructures • Cyber terrorists • States (cyber commands)

The first is technical and concerned with malware (viruses, worms, etc.) and system intrusions. The second is concerned with cyber crime and cyber espionage. The third is discourse driven and initiated by the US military, initially focusing on matters of cyber war but increasingly also on critical infrastructure protection within the realm of civil defence/protection or homeland security. Each of them is uniquely shaped and dominated by specific actors and revolves around particular 'referent objects' (that which is seen in need of protection, see Buzan et al. 1998) and threats, as summarised in Table 1.

A. Viruses, Worms and Other Bugs (Technical Discourse)

The technical discourse is focused on computer and networks disruptions caused by different types of malware. In 1988, the ARPANET – the precursor of today's Internet – had its first major network incident: the 'Morris Worm'. The worm used so many system resources that large parts of the early Internet went down. The rather devastating technical effect prompted the Defense Advanced Research Projects Agency (DARPA, who was in charge of the network at the time) to set up a centre to coordinate communication among computer experts during IT emergencies and to help prevent future incidents: a Computer Emergency Response Team (CERT) (Scherlis et al. 1990). This centre, now called the CERT Coordination Center, still plays a considerable role in computer security and served as a role model for similar centres around the world. Around the same time, the anti-virus industry emerged, bringing with it techniques and programs for virus recognition, destruction and prevention.

The worm also had a substantial psychological and, subsequent, political impact by making policy-makers aware of the Internet's insecurity and unreliability. While it was acceptable in the 1960s for pioneering computer professionals to hack and investigate computer systems,

the situation changed by the 1980s. Society had become dependent on computing for everyday business practices and other basic functions. Tampering with computers suddenly meant potentially endangering people's careers and property; and some even said their lives (Spafford 1989). Ever since, malware, as 'visible' proof of the persuasive insecurity of the information infrastructure, has remained in the limelight of the cyber security discourse – and provides the back-story for the other two discourses.

B. Cyber Crooks and Digital Spies (Crime-Espionage Discourse)

The cyber crime and technical discourses, respectively, are very closely related. The development of IT law (and, more specifically, Internet or cyber law) in different countries plays a crucial role in the second discourse, largely as it allows the definition and prosecution of a misdemeanour. Not surprisingly, the development of legal tools to prosecute unauthorized entry into computer systems (like the Computer Fraud and Abuse Act of 1986 in the United States) coincided with the first serious network incidents (cf. Mungo and Clough 1993).

Cyber crime has come to refer to any crime that involves computers and networks, like the release of malware or spam, fraud, and many other things. Until today, notions of computer-related economic crimes determined the discussion about computer misuse. However, a distinct national-security dimension was established when computer intrusions (a criminal act) were clustered together with the more traditional and well-established espionage discourse. Prominent hacking incidents – such as the intrusions into high-level computers perpetrated by Milwaukee-based '414s' gang (6 teenagers) – led to a feeling in policy circles that there was a need for action (Ross 1991): If teenagers were able to penetrate computer networks that easily, it was assumed that better organized entities such as states would be even better equipped to do so. Over the years, this discourse has become particularly focused on so-called advanced persistent threats, a cyber attack category which connotes an attack with a high degree of sophistication and stealthiness over a prolonged duration of time. The attack objectives typically extend beyond immediate financial gain.

C. Information Warfare and Critical Infrastructures (Military-Civil Defence Discourse)

The link between information technology and national security was firmly established in military writings in the time after the Second World War (Edwards 1996). But it was the Second Persian Gulf War of 1991 that created a watershed in US military thinking about cyber war. Military strategists saw the conflict as the first of a new generation of conflicts, in which physical force alone was not sufficient, but was complimented by the ability to win the information war and to secure 'information dominance'. As a result, American military thinkers began to publish scores of books on the topic and developed doctrines that emphasized the ability to degrade or even paralyse an opponent's communications systems (cf. Campen 1992).

In the mid-1990s, the advantages of the use and dissemination of Information Communication Technology (ICT) that had fuelled the revolution in military affairs were no longer seen only as a great opportunity providing the country with an 'information edge' (Nye and Owens 1996), but were also perceived as constituting an over-proportional vulnerability vis-à-vis a malicious state and non-state actors (Rattray 2001). This perception was shaped by the larger strategic

context that emerged for the United States after the Cold War. The new environment was characterised by more dynamic geostrategic conditions, numerous areas and issues of concern as well as smaller, more agile and more diverse adversaries. As a result of the difficulties to locate and identify enemies, parts of the focus of security policies shifted away from actors, capabilities, and motivations to general vulnerabilities of the entire society. Global information networks seemed to make it much easier to attack the US asymmetrically, as such attacks no longer required big, specialized weapons systems or an army: borders, already porous in many ways in the real world, were non-existent in cyberspace. It seemed only a matter of time until those actors, likely to fail against American military power, would seek to bring the US to its knees by striking vital points fundamental to the national security and essential functioning of industrialized societies at home (Berkowitz 1997): critical infrastructures.

At the same time, the development of military doctrine for the information domain continued. For a while, information warfare – the new type of warfare in the information age – remained essentially limited to military measures in times of crisis or war. This shifted around the mid-1990s, when the activities began to be understood as actions targeting the entire information infrastructure of an adversary – political, economic, and military, throughout the continuum of operations from peace to war (Brunner and Dunn Cavelty 2009). NATO's 1999 intervention against Yugoslavia marked the first sustained use of the full-spectrum of information warfare components in combat. Much of this involved the use of propaganda and disinformation via the media (an important aspect of information warfare), but there were also website defacements, a number of DDoS-attacks[3], and (unsubstantiated) rumours that Slobodan Milosevic's bank accounts had been hacked by the US armed forces (Dunn 2002: 151). The increasing use of the Internet during the conflict gave it the distinction of being the 'first war on the Internet'.

D. Countermeasures

By the end of the 1990s, the three discourses had produced specific types of concepts and actual countermeasures on the national and the international level in accordance with their focus (see Table 2). Worldwide, the protection policies that transpired consisted of a three-pronged approach: A strong law enforcement pillar for countering cyber crime, private-public partnerships for critical infrastructure protection (Dunn Cavelty and Suter 2009), and private and public self-help for the rest of the networked infrastructures. It became a common pragmatic practice that everybody was quasi responsible for 'their own': governments protect government networks, militaries only military ones, companies protect theirs, and every individual out there is in charge of their own computer security.

However, there are some assets in the hands of the private sector considered so crucial to the functioning of society that governments take additional measures to ensure an adequate level of protection. These efforts are usually subsumed under the label of critical (information) infrastructure protection (CIIP). At the core of these practices, we find the strategy of preparation, meaning the preventive protection of critical infrastructures by technical means, namely information assurance practices (May et al. 2004), supplemented by the concept of resilience. Resilience, a concept which accepts that disruptions are inevitable, is commonly defined as the ability of a system to recover from a shock, either returning back to its original

3 Attempts to make a computer or network resource unavailable to its intended users, mostly by satu-rating the target machine with external communications requests so that it cannot respond to legitimate traffic, or responds so slowly as to be rendered effectively unavailable.

state or to a new adjusted state. Therefore, the concept promises an additional safety net against large-scale, major and unexpected events (Perelman 2007; Dunn Cavelty 2011b).

TABLE 2: SET OF COUNTERMEASURES AT THE END OF THE 1990S

	Technical	Crime-Espionage	Civil defence	Strategic-military
Basic Protection concept	Information Assurance			
National level	• CERTs	• Computer law	• Critical (information) infrastructure protection	• Cyber defence (for military networks) • Resilience
International level	• International CERTs • Information security standards	• Harmonization of law • Mutual judicial assistance procedures	• Resilience	• (Cyber arms control)

Particularly interesting about these policy solutions is the relatively small role of the state. The consequences of cyber vulnerabilities for the well-being of a nation are very high – at least in theory. Therefore, a national security connotation seemed a natural given as soon as the link to critical infrastructure was made in the third discourse. But while military documents and strategists were influential in shaping general threat perceptions and in bringing the issue of cyber threats to the attention of a broad audience, the reality of the main referent object – critical infrastructures, most of them in the hand of the private sector – and the nature of the threat made it impossible for the traditional national security bodies, especially the military, to play a larger role in countering it.

For example, high-level cyber attacks against infrastructure targets would likely be the culmination of long-term, subtle, systematic intrusions. The preparatory phase could take place over several years. When – or rather if – an intrusion is detected, it is often impossible to determine whether it was an act of vandalism, computer crime, terrorism, foreign intelligence activity, or some form of strategic military attack. The only way to determine the source, nature, and scope of the incident is to investigate it. This again might take years, rendering highly uncertain results. The military notion of striking back electronically or kinetically is therefore unusable in most cases.

In addition, cyber threats defy the well-proven concept of deterrence. Deterrence works if the one party is able to successfully convey to another that it is both capable and willing to use a set of available (often military) instruments against him if the other steps over the line. This requires an opponent that is clearly identifiable as an attacker and has to fear retaliation – which is not the case in the cyber domain because of the so-called attribution problem; the architecture of cyberspace makes it difficult to clearly determine those initially responsible for a cyber attack as well as to identify motivating factors. Attacks and exploits that seemingly benefit states might well be the work of third-party actors operating under a variety of motivations. At the same time, the challenges of clearly identifying perpetrators gives state actors convenient

'plausible deniability and the ability to officially distance themselves from attacks' (Deibert and Rohozinski 2009: 12). Blame on the basis of the 'cui bono'-logic (which translates into 'to whose benefit') on the other hand is not sufficient proof for political action (in most cases). Therefore, deterrence and retribution do not work in cyberspace and will not, unless its rules are changed in substantial ways, with highly uncertain benefits (Libicki 2009). Though fears of future cyber wars existed at the time, efforts to control the military use of computer exploitation through arms control or multilateral behavioural norms like agreements that might pertain to the development, distribution, and deployment of cyber weapons, or to their use, remained limited (Denning 2001), at least until recently.

3. THE 'STUXNETIFICATION' OF THE DEBATE

The set of practices as described above remained fairly stable for more than a decade. More recently, however, the threat perception has changed – and with it how some governments address the issue. Four recent trends and developments have solidified the impression that cyber disturbances are increasingly dangerous and aggressive and that governments should react more forcefully to them – particularly by enhancing their own offensive capabilities.

First, there is heightened concern with the rising level of professionalization coupled with the obvious criminal (or even strategic) intent behind attacks. Advanced malware is targeted: A hacker picks a victim, scopes the defences and then designs malware to get around them (Symantec 2010). The most prominent example for this kind of malware is Stuxnet (addressed below). This development goes in sync with the development of the cyber crime market, which is driven by the huge sums of money available to criminal enterprises at low risk of prosecution (Panda Security 2010).

Second, the main cyber 'enemy' has been singled-out: there is an increase in allegations that China is responsible for cyber espionage in the form of high-level penetrations of government and business computer systems in Europe, North America, and Asia. Because Chinese authorities have stated repeatedly that they consider cyber space to be a strategic domain and by mastering it they may be able to equalise the existing military imbalance between China and the US more quickly, many US officials readily accuse the Chinese government of deliberate and targeted attacks or intelligence gathering operations (Ball 2011).

Third, there is an increase in sophisticated hacktivism activities. WikiLeaks, for example, has added yet another twist to the cyber espionage discourse. Acting under the hacker-maxim 'all information should be free', this type of activism deliberately challenges the self-proclaimed power of states to keep information, which they think could endanger or damage national security, secret. Related are the multifaceted activities of hacker collectives such as Anonymous or LulzSec, who humiliate high-visibility targets by DDoS-attacks, break-ins and release of sensitive information. In addition, more and more conflicts of political or economic nature have a cyber(ed)-component these days (Deibert et al. 2012; Demchak 2010), which often includes hacktivism activities. Perhaps the most prominent example is the Estonian 'cyber war' case of 2007.

Fourth, the discovery of the computer worm Stuxnet in 2010 changed the overall tone and intensity of the debate once and for all. Stuxnet is a very complex programme. It is likely that writing it took a substantial amount of time, advanced-level programming skills and insider knowledge of industrial processes. Therefore, Stuxnet is probably the most expensive malware ever found. In addition, it behaves differently from the normal criminal-type malware: it does not steal information and it does not herd infected computers into so-called botnets to launch further attacks from (Gross 2011). Rather, it looks for a very specific target: Stuxnet was written to attack Siemens' *Supervisory Control and Data Acquisition* (SCADA) systems that are used to control and monitor industrial processes. In August 2010, the security company Symantec noted that 60% of the infected computers worldwide were in Iran. Moreover, reports alleged that the Iran nuclear program had been delayed as some centrifuges had been damaged. The picture that materializes from the pieces of this puzzle seems to suggest that only one or several nation states – the 'cui bono' logic pointing either to the US or Israel – would have the capability and interest to produce and release Stuxnet in order to sabotage the Iranian nuclear program (Farwell and Rohozinski 2011).

4. UNINTENDED SIDE-EFFECTS: CAUSES AND REMEDIES

This 'story', which is indeed convincing and plausible, has seized to be a mere story: it has become the truth, despite the fact that the evidence for Stuxnet being a government-sponsored cyber weapon directed at Iran is purely circumstantial. It may in fact never be possible to know for certain who gave the order to program Stuxnet, who actually did it, and what the intent behind it was. However, this is strangely irrelevant: The only thing that does matter in this instance is what states make of it – because it is their actions and reactions that create political reality.

The reaction is that more and more states are opening up or enhancing 'cyber commands', which are military units for cyber war activities, because just the possibility that one or several state actors are behind the computer worm means that this *could* mark the beginning of the unchecked use of cyber weapons in open or more clandestine military aggressions. Though consolidated numbers are hard to come by, the amount of money spent on defence-related aspects of cyber security is rising. The new cyber military-industrial complex, for instance, is estimated to make returns between $80-billion and $150-billion US dollars a year, with big defence companies like Boeing and Northrop Grumman repositioning themselves to service the expanding cyber security market (Deibert and Rohozinski 2011).

Following the strategic logic, several states have ramped up their rhetoric. For example, Iranian and Indian officials have gone on public record condoning hackers who work in the state's interest. The White House's new International Strategy for Cyberspace of 2011 states that the United States reserves the right to retaliate against hostile acts in cyberspace with military force. Because cyber capabilities cannot be divulged by normal intelligence gathering activities, uncertainty and mistrust are on the rise. The first signs of a 'cyber security dilemma' are discernible: Although most states still predominantly focus on cyber defence issues, measures

taken by some nations are seen by others as covert signs of aggression by others, and will likely fuel more efforts to master 'cyber weapons'.

As pointed out in the introduction, reacting this way is not inevitable (though arguably understandable). It is a matter of choice, or at least a matter of a political process that has produced this particular outcome. Unfortunately, it is making both the virtual but also the real world less and not more safe. The overall aim of cyber security policy is to reduce the risks in and through cyberspace. If certain reactions or policy approaches are becoming complicit in creating more insecurity, then they should be corrected. The good news is that there are alternatives both in framing the issue and in countering it, and that both these frames and these countermeasures are already in place, as shown above. For a reframing to become possible, however, skewed threat perceptions that are the outcome of government circles to focus too much on high-impact, low-probability events need to be corrected.

A. Why the Threat is Persistently Overrated

Every political, economic and military conflict nowadays has a cyber(ed)-component. Furthermore, criminal and espionage activities with the help of computers happen every day. It is a fact that cyber incidents are continually causing minor and occasionally major inconveniences in the form of lost intellectual property or other proprietary data, maintenance and repair, lost revenue, and increased security costs. Beyond the direct impact, badly handled cyber attacks have also damaged corporate (and government) reputations. However, in the entire history of computer networks, cyber attacks have never caused serious long-term disruptions. They are risks that can be dealt with by individual entities using standard information security measures and their overall costs remain low in comparison to other risk categories like financial risks.

Despite this, the threat keeps being 'hyped' in policy circles. There are several reasons for this: First, psychological research has shown that risk perception is highly dependent on intuition and emotions, also the perceptions of experts (Gregory and Mendelsohn 1993). Cyber risks, especially in their more extreme form, fit the risk profile of so-called 'dread risks', which appear uncontrollable, catastrophic, fatal, unknown and basically uncontrollable. There is a propensity to be disproportionally afraid of these risks despite their low probability, which translates into pressure for regulatory action of all sorts and willingness to bear high costs of uncertain benefit.

Second, combating cyber threats has become a highly politicised issue. Therefore, official statements about the threat must also be seen in the context of different bureaucratic entities that compete against each other for resources and influence or of politicians taking up this new and politically 'hot' issue. This is usually done by stating an urgent need for action and describing the overall threat as big and rising. Furthermore, being a cyber-expert has become a lucrative market, but only if the problem is continuously portrayed as grave.

Third, the media loves the idea of cyber-'anything' in connection with disaster, and routinely features sensationalist headlines that cannot serve as a measure of the problem's scope. By reporting only on a certain type of cyber-issue, they distort the threat perception. Some IT security companies have recently warned against overemphasizing so called advanced persistent threat attacks just because we hear more about them (Verizon 2010: 16). Only about 3% of all

incidents in 2010 were considered so sophisticated that they were impossible to stop. The vast majority of attackers go after low hanging fruit, which are small to medium sized enterprises with bad defences and little security awareness (Maillart and Sornette 2010).

B. From Vulnerability Assumptions to Threat Assessments

Since the effects of cyber attacks are potentially devastating, the temptation to not only think about worst-case scenarios but also give them a lot of (or rather too much) weight, despite their low probability, is high. This problem is aggravated by a broader tendency in security politics. The handling of issues is directly linked to level of knowledge, but more importantly non-knowledge about threats. Traditional threat analysis looked at the capability or potential of enemies and their intent or motivation, in addition to one's own vulnerability. Cyber threats, however, are highly diffuse and many aspects are unknowable. There is no reliable data for loss or damage estimation within our current cyber pattern of cyber usage and it is very unlikely that there will ever be satisfactory solutions to this data problem. Attempts to collect it have failed due to insurmountable difficulties in establishing what to measure, how to measure it, and what to do about incidents that are discovered very late, or not at all (Sommer and Brown 2011: 12).

Missing knowledge of this sort has led to increasing use of vulnerability-based analysis, based solely on the identification of weaknesses (Jenkins 2006: 120). When looking at vulnerabilities, the follow-up question is: 'what could go wrong?' and the natural answer is: 'everything'. This almost automatically leads to worst-case scenarios. However, these scenarios have a habit to become reified in the political process. When this happens, they are turned into real threats, not potentials, based not on knowledge about the intentions and capabilities of potential adversaries but mainly on policy-makers' fears (Furedi 2008: 652).

Such thinking distracts attention from the highly relevant questions of 'what can' and 'what is likely' to happen (Furedi 2008: 653). The correct assumption that modern societies and their armed forces depend on the smooth functioning of information and communication technology does not automatically mean that this dependence or vulnerability *will* be exploited. Patching all the vulnerabilities of modern societies is outright impossible and also not politically or economically desirable. Therefore, the policy community must return to level-headed threat assessments that ask 'who has the interest and the capability to attack us and why would they?'

At the moment, most experts agree that strategic cyber war (and catastrophic attacks) remains highly unlikely in the foreseeable future, mainly due the uncertain results such a war would bring, the lack of motivation on the part of the possible combatants, and their shared inability to defend against counterattacks (Sommer and Brown 2011). Cyber crime and cyber espionage, both political and economic, are a different story: they are here and will remain the biggest cyber risks in the future. Very clearly, they deserve the full attention of the policy community much more than their unlikely counterparts.

5. CONCLUSION

Thinking about (and planning for) worst-case scenarios is a legitimate task of the national security apparatus. However, catastrophic incidents should never receive too much attention at the expense of more plausible and possible cyber problems. Using too many resources for high impact, low probability events – and therefore having less resources for the low to middle impact and high probability events – does not make sense, neither politically, nor strategically and certainly not when applying a cost-benefit logic.

Despite the increasing attention cyber security is getting in security politics, computer network vulnerabilities are mainly a business and espionage problem. Further militarising cyberspace based on the fear of other states' cyber capabilities is pointless. While it is undisputed that the cyber dimension will play a substantial role in future conflicts of all grades and shades, threat-representations must remain well informed and well balanced at all times in order to rule out policy (over)reactions with too high costs and uncertain benefits. Regardless of how high we judge the risk of a large-scale cyber attack, military-type countermeasures will not be able to play a substantial role in cyber security because of the nature of the attacker, the nature of the attacked, and the nature of the cyber(ed)-environment. Investing too much time talking about them or spending increasing amounts of money on them is not going to make cyberspace more secure – quite the contrary.

Cyberspace is only in parts controlled or controllable by state actors. At least in the case of democracies, power in this domain is in the hands of private actors, especially the business sector. Much of the expertise and many of the resources required for taking better protective measures are located outside governments. The military – or any other state entity for that matter – does not own critical (information) infrastructures and has no direct access to them. Protecting them as a military mandate is an impossibility and considering cyberspace as an occupation zone is an illusion. Militaries cannot defend the cyber space of their country – it is no space where troops and tanks can be deployed because the logic of national boundaries does not apply.

Undoubtedly, however, attacks on information technology, manipulation of information, or espionage can have serious effects on the present and/or future of defensive or offensive effectiveness of one's own armed forces. First and foremost, militaries should therefore focus on the protection and resilience of their information infrastructure and networks, particularly the critical parts of it, at all times. Beyond this, governments and military actors should acknowledge that their role in cyber security can only be a limited one, even if they consider cyber attacks to be a major national security threat. Cyber security is and will remain a shared responsibility between public and private actors. Governments should maintain their role in protecting critical infrastructure where necessary, while determining how to best encourage market forces to improve the security and resilience of company owned networks.

REFERENCES

Ball, D. (2011), 'China's Cyber Warfare Capabilities', *Security Challenges*, 7/2: 81–103.

Berkowitz, B.D. (1997), 'Warfare in the Information Age', in Arquilla, J. and Ronfeldt, D.F. (eds.), *In Athena's Camp: Preparing for Conflict in the Information Age* (Santa Monica: RAND), 175–90.

Brunner, E. and Dunn Cavelty, M. (2009), 'The Formation of In-Formation by the US Military', *Cambridge Review of International Affairs*, 22/4: 629–646.

Buzan, B., Wæver, O. and de Wilde, J. (1998), *Security: A New Framework for Analysis* (Boulder: Lynne Rienner).

Campen, A.D. (1992) (ed.), *The First Information War: The Story of Communications, Computers and Intelligence Systems in the Persian Gulf War* (Fairfax: AFCEA International Press).

Deibert, R. and Rohozinski, R. (2011), 'The new cyber military-industrial complex', The Globe and Mail, March 28, 2011. Accessed 3 March 2012, http://www.theglobeandmail.com/news/opinions/opinion/the-new-cyber-military-industrial-complex/article1957159/.

Deibert, R. and Rohozinski, R. (2009), 'Tracking GhostNet: Investigating a Cyber Espionage Network', Information Warfare Monitor (Toronto: The Munk School of Global Affairs).

Deibert, Ronald J., Rohozinski, R. and Crete-Nishihata, M. (2012), 'Cyclones in cyberspace : informatoin shaping and denial in the 2008 Russia-Georgia war', Security Dialogue 43/3: 3-24.

Demchak, C. (2010), 'Cybered Conflict as a New Frontier', *New Atlanticist,*. Accessed 3 March 2012, http://www.acus.org/new_atlanticist/cybered-conflict-new-frontier.

Denning, D. (2001), 'Obstacles and Options for Cyber Arms Controls', paper presented at the Arms Control in Cyberspace Conference, Heinrich Böll Foundation, Berlin, 29-30 June 2001. Accessed 3 March 2012 http://www.cs.georgetown.edu/~denning/infosec/berlin.doc.

Dunn, M. (2002), *Information Age Conflicts: A Study of the Information Revolution and a Changing International Operating Environment* (Zurich: Center for Security Studies).

Dunn Cavelty, M. (2008), *Cyber-Security and Threat Politics: US Efforts to Secure the Information Age* (London: Routledge).

Dunn Cavelty, M. (2011a), 'Cyber-Allies: Strengths and Weaknesses of NATO's Cyberdefense Posture', *IP Global Edition*, 12/3: 11–15.

Dunn Cavelty, M. (2011b), 'Systemic cyber/in/security – From risk to uncertainty management in the digital realm', *Swiss Re Risk Dialogue Magazine*, 15 September.

Dunn Cavelty, M. and Suter, M. (2009), 'Public-Private Partnerships are no Silver Bulled: An Expanded Governance Model For Critical Infrastructure Protection', *International Journal of Critical Infrastructure Protection*, 2/4: 179–87.

Edwards, P.N. (1996), *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, MA: MIT Press).

Farwell, J.P. and Rohozinski, R. (2011), 'Stuxnet and the Future of Cyber War', *Survival: Global Politics and Strategy*, 53/1: 23–40.

Furedi, F. (2008), Fear and Security: A Vulnerability-led Policy Response. Social Policy & Administration, 42: 645–661.

Gregory, R. and Mendelsohn, R. (1993), 'Perceived Risk, Dread, and Benefits', *Risk Analysis* 13/3: 259–64.

Gross, M.J. (2011), 'Stuxnet Worm: A Declaration of Cyber-War', *Vanity Fair*, April.

Jenkins, M. J. (2006), 'The new age of terrorism', in: Kamien, D (ed.), *McGraw-Hill Homeland Security Handbook* (New York: McGraw-Hill).

Libicki, M.C. (2009), *Cyberdeterrence and Cyberwar* (Santa Monica: RAND).

Maillart, T. and Sornette, D. (2010), 'Heavy-Tailed Distribution of Cyber-Risks', *The European Physical Journal* B, 75/3: 357–64.

May, Chris et al. (2004), 'Advanced Information Assurance Handbook', CERT®/CC Training and Education Center, CMU/SEI-2004-HB-001 (Pittsburgh: Carnegie Mellon University).

Mungo, P. and Clough, B. (1993), *Approaching Zero: The Extraordinary Underworld of Hackers, Phreakers, Virus Writers, and Keyboard Criminals* (New York: Random House).

Nye, J.S. Jr. and Owens, W.A. (1996), 'America's Information Edge', *Foreign Affairs*, March/April: 20–36.

Panda Security (2010), *Panda Security Report: The Cyber-crime Black Market*: Uncovered (Bilbao).

Perelman, L.J. (2007), 'Shifting Security Paradigms: Toward Resilience', in J.A. McCarthy (ed.), 'Critical Thinking: Moving from Infrastructure Protection to Infrastructure Resilience', *CIP Program Discussion Paper Series* (Washington: George Mason University), 23–48.

Rattray, G. (2001), *Strategic Warfare in Cyberspace* (Cambridge, MA: MIT Press).

Ross, A. (1991), 'Hacking Away at the Counterculture', in C. Penley and A. Ross (eds.), *Technoculture* (Minneapolis: University of Minnesotta Press), 107–34.

Scherlis, W.L., Squires, S.L. and Pethia, R.D. (1990), 'Computer Emergency Response,' in P. Denning (ed.), *Computers Under Attack: Intruders, Worms, and Viruses* (Reading: Addison-Wesley), 495–504.

Sommer, P. and Brown, I. (2011), *Reducing Systemic Cyber Security Risk*, Report of the International Futures Project, IFP/WKP/FGS(2011)3 (Paris: OECD).

Spafford, E.H. (1989), 'The Internet Worm: Crisis and Aftermath', *Communications of the ACM*, 32/6: 678–87.

Symantec (2010), Internet Security Threat Report, Vol. 16 (Mountain View).

Verizon (2010), 2010 Data Breach Investigations Report: A Study Conducted by the Verizon RISK Team in cooperation with the United States Secret Service (New York).

Chapter 3
Cyber Conflict – Theory & Principles

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Beyond Domains, Beyond Commons: Context and Theory of Conflict in Cyberspace

Jeffrey L. Caton
Science and Technology Division,
Center for Strategic Leadership
U.S. Army War College
Carlisle Barracks, Pennsylvania, U.S.A
jeffrey.caton@us.army.mil

Abstract: This paper examines implications of the collective cognitive blind spot of national security leaders with regard to conflict and warfare in and through cyberspace. It argues that the view of cyberspace as a contested domain within a global commons is not sufficient to address the full range of conflict therein. It posits that deliberate examination of the ontology and evolution of cyberspace is essential to properly inform the management of resources, forces, and risk. It discusses analytical frameworks to explore the fundamental structures of cyberspace and endeavors to provide the theoretical underpinning necessary to inform the broader dialogue addressing concepts such as complexity and emergence, self-organization and self-governance, human-machine integration, influences of ethics and philosophy, and the blurring of distinction between the cognitive, content, and connectivity dimensions. It strongly encourages leaders in cyberspace security planning to adopt a bifurcated approach that not only addresses the immediate challenges in cyberspace, but also includes a parallel and distinct effort to examine and characterize future manifestations of cyberspace.

Keywords: *cyberspace, theory, ontology, evolution, conflict, commons*

1. INTRODUCTION

Often, cyberspace security resembles the analogy of the blind men and the elephant—that is, the scope of activity reflects only the part of cyberspace encountered. Working together, the blind men can divine the whole of the elephant from their perceived parts. Unlike the constant form of the elephant, cyberspace is changing rapidly; this creates an expanding chasm between the perceived and the actual cyberspace environment. This situation may degrade a nation's ability to conduct critical analysis regarding future investments in cyberspace infrastructure,

The views expressed are those of the author and do not necessarily reflect official policy or position of the United States Army War College, the Department of the Army, the Department of Defense, or any other Department or Agency within the U.S. Government. All information and sources for this paper were drawn from unclassified material.

personnel, and education. This paper argues that the current demands of international security entail consideration of contexts beyond the limited view of cyberspace as a contested domain. Further, understanding the future strategic security environment requires the examination of cyberspace ontology, a study of its evolution, and the development of theory regarding activities in cyberspace. To investigate this assertion, we first examine the model of cyberspace as a contested domain and then broaden to view cyberspace as a commons encompassing all elements of national power. Next, we explore the complex structure and dynamic nature of the commons and the related international security implications. Finally, we reach beyond the commons view to address the ontology and future of cyberspace itself. The goal is to broaden the reader's perspective regarding the context and theory of cyberspace in the current global security environment as well as to encourage understanding of the fundamental nature of cyberspace and its complex and dynamic evolution beyond the complacency of technical stovepipes.

2. CYBERSPACE AS A CONTESTED DOMAIN

What are the current perceptions of the roles of cyberspace in the international security environment? The U.S. government defines cyberspace as "the interdependent network of information technology infrastructures, and includes the Internet, telecommunications networks, computer systems, and embedded processors and controllers" and it "also refers to the virtual environment of information and interactions between people" [1]. Considering the well-documented historic cases of Titan Rain (2004) [2], Estonia (2007) [3], Georgia (2008) [3], and Operation Buckshot Yankee (2008) [4] as well as recent mysteries such as the Conficker and Stuxnet worms, cyberspace can be portrayed as a contested domain. Consistent with this view, U.S. Cyber Command achieved Full Operational Capability in October 2010, with its mission to direct operations and defense of Department of Defense (DoD) networks, conduct full-spectrum military cyberspace operations, and ensure U.S. and Allied freedom of action in cyberspace and deny the same to adversaries [5]. Complementary to this mission, the *DoD Strategy for Operating in Cyberspace* was released in July 2011 as the first DoD unified strategy for cyberspace [6]. These DoD initiatives mesh well with the tenets of the June 2010 North Atlantic Treaty Organization (NATO) Policy on Cyber Defence which emphasizes prevention, resilience, and non-duplication. This policy strives to "integrate cyber defence considerations into NATO structures and planning processes in order to perform NATO's core tasks of collective defence and crisis management" [7]. The collective approach of these 28 nations is to treat cyberspace as another domain—like land, sea, air, or space—used to control and exploit with the intent of exerting influence on the other domains. The immediate focus is on defense; although clearly a perfect defense is not possible even within the limits of a military domain. It becomes even more challenging when government protection systems, such as the EINSTEIN 3 intrusion-detection system, are considered to protect private critical infrastructure networks [8]. The strategies address recent historical events—distributed denial of service, botnets, patriotic hackers—that reflect the brute-force approach to cyberspace aggression.

Consistent with current U.S. joint force doctrine, cyberspace operations encompass the three dimensions of the information environment—cognition, content, and connectivity. At present, the content and connectivity portions are emphasized since they involve the software and

hardware portions of cyberspace. However, many existing plans focus on threats and activities that are actually historical in terms of the relative rate of cyberspace growth and change. Without a more progressive context, near-term activities in cyberspace may be misguided, and long-term planning for investment and force structure there may be obsolete before they are enacted.

3. BEYOND DOMAINS TO THE COMMONS

To gain a broader outlook of cyberspace, let us consider the commons paradigm and then examine theoretical models and the related complex and dynamic nature. Commons are areas not controlled or owned by any single entity, and states and non-state actors use them to conduct commerce and communication [9]. In such a commons, a nation could exercise cyberpower as its "ability to use cyberspace to create advantages and influence events in all the operational environments and across the instruments of power" [10]. The cyberspace commons has unique features that facilitate use to entities smaller than nation-states. There are many low-cost options available that provide users with reliable access. Users' persona in cyberspace need not match their true appearance and it is possible to have multiple representations simultaneously; this degree of anonymity may challenge efforts to attribute activities in cyberspace. Cyberspace can enable one to initiate a variety of physical effects across vast distances at almost instantaneous speeds [11]. Thus, determining if cyberspace should be treated as a domain or a commons depends on the level of application. Developing force structures and unit competencies may be served best by the domain view; examining the extent and priority of cooperative engagement amongst countries may be served best by the commons view.

A. Theoretical Models

What are some methodologies that provide a context beyond the domain concept for evaluating activity in cyberspace? Consider two examples of theoretical frameworks to guide strategy development and implementation.

1. Ecosystem Model

In March 2011, a Department of Homeland Security paper proposed a "healthy cyber ecosystem" as a model for enabling security in cyberspace. Based in part on the human immune system, the concept envisions future cyber devices with "innate capabilities that enable them to work together to anticipate and prevent cyber attacks, limit the spread of attacks across participating devices, minimize the consequences of attacks, and recover to a trusted state" [12]. The model uses mutually supporting healthy cyberspace devices working together proactively and dynamically to assess the severity of any "infection" and respond when an appropriate "alert threshold" is exceeded. But anomalous and negative activities in cyberspace are not always attacks; they may be manifestations of complex interactions and emergence of unanticipated behavior. Thus, automated active defense systems must ensure that attacks are clearly distinguished from mere alerts [12]. Such measures may create an information environment that is largely self-regulating with respect to mundane threats. However, just as strengthening human immune systems may have the counterproductive effect of producing more virulent strains of germs, one may ponder what new threats could emerge in a healthy cyberspace environment.

2.Naval Theory Analogy

Traditional (i.e., Mahanian) naval theory offers value for modeling military activities as domain operations within a global commons. When one connects major ports in the littoral area ("brown water") to other ports in the world, "sea lines of communication" emerge in the broad ocean ("blue water") that have strategic importance based on factors such as geography and traffic volume [13]. Similarly, one can map cyberspace to show "cyber lines of communication" and nodes with tactical, operational, and strategic implications, perhaps even choke points—the "blue water cyberspace" equivalent of the Strait of Hormuz.

Barney [14] uses the 1982 United Nations Convention on the Law of the Sea framework to examine issues related to routing information through cyberspace. He proposed *national cyberspace* as "the region of Cyberspace in which individual States require substantial sovereign rights to preserve the political and economic security." He divided this region into *internal cyberspace*, "where a State may exercise complete sovereignty," and *territorial cyberspace* "through which, and to which, governments, commercial enterprises, or private organization allow generally unrestricted access." *International cyberspace* is a region with no physical analogy to international waters; it "is not a physical place; it is a *characteristic* of Cyberspace by which a data packet is not physically present anywhere but is merely in transit" [14].

Like ship traffic crossing oceans, consider information traffic as packages of data moving across electromagnetic waves in cyberspace. The right of innocent passage provides "the right to traverse the territorial sea in a continuous and expeditious manner, so long as the passage is not prejudicial to the peace, good order, or security of the coastal State." The right of transit passage provides "freedom of navigation and overflight solely for the purpose of continuous and expeditious transit of the international strait between one part of the high seas or an exclusive economic zone and another part." Transit passage provides the advantages that "forces may transit in their normal mode of operation (i.e., warfighting) and bordering States may not suspend the right of transit passage through international straits." Applying these principles to cyberspace, Barney concludes that computer network attack (CNA) "may be lawfully transmitted through the international telecommunications infrastructure, including Internet routers physically located in neutral States." He notes in his scenario that such passage does not violate territorial sovereignty, nor comprise an act of force in the intermediate territory, nor violate the status of neutral States [14]. Perhaps this framework could allow data packets to be "nationally flagged" akin to ships—thus having data itself represent sovereignty and facilitate development of diplomatic measures that allow (or deny) packet transit. In turn, this could help the international community respond to unauthorized rerouting of packets via router disruption, such as China is accused of doing in 2010 [15].

This model may also facilitate development of theory related to virtual environments—that is, the immersion of the cognitive mind in non-physical landscapes defined by code often distributed among many machines. Perhaps the concept of naval subsurface activity can model virtual activity. Arguably, it is more difficult to "track" individuals, groups, and activities when they go below the "normal surface" of cyberspace (whatever that truly is) into the multi-dimensional subsurface virtual environment. Conversely, when they "re-surface," their presence may be more readily acknowledged and attributed. The concept of riverine operations—those that focus on a nation's inland waters—may offer models for devolved operations in cyberspace "backwaters."

This would focus on activities that use older technology, such as telephone modems and DOS-based bulletin boards, instead of modern Internet connections [16]. Regardless of how packets travel, their movement through cyberspace may meet resistance in the same way ships navigate through waves and currents. Thus, ensuring freedom of navigation in cyberspace must necessarily include not only adversarial efforts to deny or disrupt, but also entropic effects and physical environment impacts (e.g., power outages, solar storms).

B. Key Characteristics of its Complex Structure

The limits of cyberspace are uncharted with rapidly expanding boundaries and increasingly complicated internal configurations. With land, sea, air, or space, technology dictates the ability to access and maneuver within the commons, but the physical structure remains relatively constant. For cyberspace, technology not only dictates how we access the common, but it also empowers the manifestation of the common. Cyberspace may be unique in that access and creation are almost synonymous—that is, the technology used to access cyberspace (e.g., computers, mobile devices) becomes an inherent part of the domain. When entering cyberspace, one must consider the *reciprocity of connectivity* associated with the access. If a user connects a device to the cyberspace commons, then the whole of the commons can connect to that device—this axiom helps define the realm of the possible. Thus, it is impossible to open a perfect one-way portal into cyberspace; any data sent or accessed over cyberspace can be viewed by anyone in cyberspace. Some applications of this process go unnoticed by even experienced users, such as the demonstrated vulnerability of modern automobile electronic control units to access and command by wireless systems [17]. Granted, such access may require illegal or unethical activity, but this does not make the action impossible.

Similar to other mediums, the active cyberspace environment has discernable structure that supports the disorderly movement of its contents. Classic thermodynamic modeling characterizes such random motion and disorder as "entropy." The transmission of data over the Internet may result in its division into many subpackets sent over different paths through an unknown number and type of processors and switches. Cyberspace is inherently complex and disorderly; the degree of which only increases as cyberspace expands with additional devices and infrastructure. It is not logical to expect order to arise spontaneously out of such a cacophony. This means that deliberate energy is required to accomplish specific tasks, and designers of content and connectivity may attempt to decrease the entropy involved with their specific function. Conversely, one could consider the overlay of security measures as purposely adding entropy to a system to thwart unauthorized use (e.g., using encryption to leverage the disorder for security's advantage). To overcome such entropy, an adversary must expend effort.

Operations in cyberspace are more difficult to accurately characterize in the realm of human cognition than traditional kinetic domain operations. Thus, promises that cyberspace offers the ultimate form of achieving precision effects may be hollow. Achieving well-characterized and bounded effects in cyberspace is more difficult than doing so in physical space, and the potential for unanticipated consequences is more likely. Further, the means and methods used to achieve precision have limited utility since their design is based on a configuration of cyberspace that is destined to change (by design or coincidence). For example, the unpredictable interaction of well-designed trading algorithms led to the "flash crash" disorder in the U.S. stock market on May 6, 2010 [18].

C. Challenges of its Dynamic Nature

Internet-enabled communications have progressed to where most users consider them direct and instantaneous. This illusion reflects the shortcoming of human perception—electrons traveling at the speed of light can circle Earth in about 130 milliseconds, one-third the time of a human eye blink. In the physical world, if one fires a weapon at a 10-meter target, the bullet follows a largely predictable path based on the gross properties of the air (e.g., temperature, wind). Interactions at the molecular level are negligible compared to the bullet's momentum, thus it follows a direct path and achieves kinetic effects at the point of impact. In cyberspace, the transmission of a data packet is assumed to follow a direct path in a stable environment. In reality, one could argue that the configuration of cyberspace at the micro level may change significantly in the milliseconds it takes to press the "Enter" key. Although the effects and path appear to be direct from a human perspective, in the relative framework of cyberspace operations, they are indirect, inefficient, and slow to manifest. Thus, projecting a guaranteed path in cyberspace is nearly impossible, just as it is impossible to align the molecules of air to accommodate a passing bullet. Consider that the May 6, 2010 stock market "flash crash" was preceded by over 10,000 ultrafast-duration crashes and spikes (less than 950 ms each) over 5 years, and that these ultrafast events continue to occur [19]. Then, the goal of the cyberspace operator is to determine not only the gross properties affecting cyberspace operations (if indeed they exist) but also the potential anomalies that may arise spontaneously as well as how to operate in them.

How should nations balance the command and control of cyberspaces forces at the battlespace level with those at the strategic level? Tyugu [20] argues that application of artificial intelligence methods (e.g., neural nets, expert systems, intelligent agents) is unavoidable for such large-scale cyber defences. Leaders must consider the scope of operational effects within the commons to coordinate them with other commanders and allies as well as affected public and industry. For example, it may be prudent to receive "cyberspace over flight permission" for offensive actions that may transit the territorial cyberspace of other nations.

4. CURRENT INTERNATIONAL SECURITY IMPLICATIONS

For alliances such as NATO, what implications arise from adopting an analysis framework for international security of a contested cyberspace domain within the larger commons? Let us examine three topics—the development of cyberspace policy and strategy; operational planning considerations to address immediate issues; and future planning. These topics mirror the NATO overarching cyber defence principles of prevention, resilience, and non-duplication.

In an ideal world, a policy that maintains the cyberspace commons as a sanctuary free of conflict is laudable. However, previous and ongoing aggression in cyberspace mandates that portions of cyberspace be militarized. The November 2010 Lisbon Summit's new NATO Strategic Concept calls for a "full range of capabilities necessary to deter and defend against any threat" among which is the requirement to "develop further our ability to prevent, detect, defend against and recover from cyber-attacks, including by using the NATO planning process to enhance and coordinate national cyber-defence capabilities" [21].

Developing cyberspace deterrence is a challenging task still in its infancy. Traditional Cold War deterrence experience may have limited application in cyberspace, given the capabilities of nonstate actors as well as the possibility of cyberattacks originating from co-opted servers in neutral countries[4]. The May 2011 U.S. *International Strategy for Cyberspace* [22] includes a deterrence policy with application to all national interests (including NATO obligations). Based on the principle that "consistent with the United Nations Charter, states have an inherent right to self-defense that may be triggered by certain aggressive acts in cyberspace," the strategy states that "when warranted, the United States will respond to hostile acts in cyberspace as we would to any other threat to our country." Further, the U.S. will "reserve the right to use all necessary means—diplomatic, informational, military, and economic—as appropriate and consistent with applicable international law, in order to defend our Nation, our allies, our partners, and our interests." These words send a serious message to potential adversaries without limiting the type of U.S. response.

But if deterrence fails, what analysis can support assessment of cyber incidents to determine if they require a NATO response? There is no internationally accepted definition of when hostile actions in cyberspace are recognized as attacks, let alone acts of war. An analytical framework developed by Schmitt [23] attempts to determine if a cyber attack equates to the use of force per terms of the U.N. charter. His analysis considers the intensity of damage in seven areas (severity, immediacy, directness, invasiveness, measurability, presumptive legitimacy, and responsibility) to provide a composite assessment of the effects of the cyber attack. Further, it addresses implications of applying Article 51 of the United Nations (U.N.) Charter for attacks prior to acknowledged armed conflict, and the law of armed conflict (LOAC) criteria for acknowledged conflict.

When it is clear that aggressive actions in cyberspace require a response, alliance actions need appropriate planning and execution. Analyzing courses of actions should include thorough evaluation of nth-order effects, not only for desired military outcomes but also for related diplomatic and economic consequences [24]. While some measures of automated defense may be necessary to protect critical functions, perhaps like neural reflexes to protect one's hand from a hot stove, the indiscriminate application may create more problems than it solves. Determining the appropriate collective response is a balancing act that may require the rapid synthesis of multiple distributed systems as well as a clear representation of any automated response, perhaps aided by graphical means, to assess the impact of the countermeasures on network resources [25].

Credible responses also depend on having properly equipped and trained personnel. Can our current understanding of cyberspace properly inform resource and force management decisions? Despite some technical promises, it is risky to consider the current application of cyberspace operations as guaranteed successes. Developing theoretical principles for cyberspace may help to explore the opportunities as well as shortcomings that affect such operations. In a technical sense, planners and decision makers need to recognize that applications may be neither precise nor free from disorder. Thus, they should conduct critical, perhaps even skeptical, reviews of promised capabilities—especially when making resourcing decisions. Investments in technology to enhance effective command and control of alliance forces must consider the implications of

operating at ever-faster network speed, such as the risks associated with selecting between proactive and reactive actions to address simultaneous system challenges [26]. Changes in the cyberspace domain and commons may be significant in the time between decision for action and its execution. Also, responsible offensive actions should incorporate "cyberspace battle damage assessment" to help ensure the necessity, distinction, and proportionality of the effects meet acceptable norms. The access to nearly limitless data may be used wisely for evaluation, but biases in data mining methodologies may reinforce and propagate cognitive blind spots. The best way to avoid such pitfalls is to examine objectively and holistically the fundamental nature of cyberspace and to envision its evolution and future embodiment.

5. BEYOND THE COMMONS – ONTOLOGY AND FUTURE

Are we developing methods to achieve situational awareness at all levels of cyberspace? This section discusses the ontology of cyberspace and recommends action in four areas to facilitate international security efforts: develop cyberspace theory; assess cyberpower of global actors; anticipate radical change; and bifurcate future-focused efforts from current operational activities.

Military activities in other domains (land, sea, air) often strive to gain local and temporal control of such domains. In contrast, cyberspace can be considered an artificial domain created for the purpose of exercising control or governing activities. It requires energy to exist (e.g., use of the electromagnetic spectrum) and its control can be first-order—conscious and deliberate—or various levels of nth-orders that may be unconscious, accidental, or emergent. It exists as part of a larger commons, both physical and virtual. To prepare for conflict beyond our current technological manifestation of cyberspace, even the commons model is insufficient. For a truly holistic view, one must examine the ontology of cyberspace (i.e., its fundamental essence) and determine how its current form fits into its overarching evolutionary path. Cyberspace ontology must address fundamental issues, such as the balance of dynamic stability for activity in cyberspace; the self-organization and self-regulation of its functions; the modeling of entropy to include concepts of convergence, divergence, and emergence; and the changes in the cognitive dimension caused by more sophisticated human-machine interfaces (e.g., neuralprosthetics [27]). Proper cyberspace theory can provide the foundation necessary to explore these ontological themes.

Starr [28] advocates that proper cyberspace theory address five areas: definition of terms; categorization and structure of theory elements; explanation of elements by analysis and example; connection of elements for comprehensive examination; and anticipation of future activities. As cyberspace theory is refined, it should be used to assess the relative strength of global actors possible through cyberpower. Future embodiments of cyberspace will likely follow the model of human conflict described by the Clausewitzian trinity of emotion, reason, and chance. As witnessed in the so-called "Arab Spring" events, social media via cyberspace can provide a conduit for human expression to force change on the world stage [29]. Pursuing holistic situational awareness can help decision makers distinguish aggression in cyberspace

from coincidental events with negative repercussions. This may be crucial during times of increased global tension. It is unlikely that cyberspace will lift the fog of war or make the application of force less subject to chance; entropy and emergence simply cannot be quantified in all circumstances. Cyberspace strategies that anticipate flaws and failures, and emphasize resilience by design, may provide enduring principles for the future.

With its increasingly complex and dynamic nature, future embodiments of cyberspace may exhibit radical change. If its structure progresses toward self-organization and self-regulation, cyberspace may surpass fully human design and control. Important research is addressing some specific aspects of change, such as the behavior and long-term strategic evolution of botnet armies [30]. Sornette [31] examined how the strengths of heterogeneity and coupling interactions among systems may shift their overall behavior from synchronization to self-organization. Of note is that extreme-risk events may occur more often than predicted for systems with low heterogeneity and high coupling—basically the situation one might find in centralized network controls with standardized desktops.

We must refine theoretical models to reflect how the balance shifts among the cognitive, content, and connectivity dimensions in the information environment. This may ameliorate the current overemphasis on information technology, as its influence may diminish. Leaders should anticipate significant blurring of the cognitive and connectivity dimensions as human-machine interfaces become more engrained and pervasive. Temmingh and Geers [32] have examined some the present challenges of distinguishing real individuals from potentially multiple cyber persona. For the future, leaders should consider the possibilities presented by blurring of the cognitive and content dimensions as information is consolidated in cloud-type applications and the collective computational and memory capacities of machines exceed that of humankind. Koch and Hepp [33] explore the possible roles of quantum mechanics in creating higher brain functions (e.g., perception, consciousness, free will). Eventually, the examination of cognition will expand to the broader human condition to include concepts of morality and ethics, and perhaps theology. This presents an essential question: is cyberspace merely a manifestation of technology, or possibly a fundamental step in human evolution?

Consider two concepts regarding the potential role of cyberspace in human evolution. First is the concept of "the singularity" explored by futurist Ray Kurzweil [34] as a "future period during which the pace of technological change will be so rapid, its impacts so deep, that human life will be irreversibly transformed." Kurzweil posits three overlapping revolutions surrounding this event—genetics, as an intersection of information and biology; nanotechnology, as an intersection of information and the physical world; and robotics, as a growth of artificial intelligence. The second concept is "the noosphere" explored by Pierre Teilhard de Chardin [35] as an extension of the biosphere to the realm of human thought. In *The Phenomenon of Man*, he describes it as "much more coherent and just as extensive as any preceding layer, it is really a new layer, the 'thinking layer,'…outside and above the biosphere there is the noosphere." Teilhard de Chardin posited some implications for humanity like those of Kurzweil, albeit through an approach of philosophy vice technology. Whether such predictions come to fruition is not the point; their ideas influence the views of human behavior that in turn influence activities of creation and utilization in cyberspace. Since it is doubtful that legal and governance regimes

will keep pace with a dynamic cyberspace environment, the establishment of a cooperative set of cyberspace ethics and value may facilitate stability and organization.

The urgent needs of international security leave few resources available for the study of cyberspace ontology and future. Relegating such activity to a mere afterthought of domain operations promises its failure, or at best, an empty victory. Thus, a bifurcated approach to policy, strategy, planning, and military preparation can best serve international cyberspace security. The first part—addressing immediate challenges in cyberspace—is in place, albeit with limitations. Actions are often reactive and ad hoc, with a decision-making context that may lag technology and not consider synergistic implications. The second part—examining future manifestations of cyberspace—can provide the cognitive foundation that informs the development of strategy, doctrine, force structure, and prioritization of resources; these in turn can help achieve unity of effort amongst all instruments of national power. This must be a separately resourced effort focused on development of theory as well as the study of ontology and evolution. This may require bold leadership and perhaps the courage to risk considering concepts that may appear foolish at times.

6. SUMMARY

While we can evaluate certain aspects of international competition and conflict in cyberspace using a domain model, the proper examination requires a holistic approach that includes concepts of the commons as well as a conscious future-directed model that recognizes the continuing evolution of cyberspace. Cyberspace theory should embrace the potential for radical emergent behavior that may shift the balance of influence among the cognition, content, and connectivity dimensions. This requires the deliberate and thoughtful pursuit of cyberspace theory as a continuing dialogue that may include multiple frameworks for analysis. This may not occur without a formal bifurcated approach to international efforts—one that is integrated to address both the pragmatic and urgent present challenges as well as a separately resourced effort dedicated to examining the changing nature of cyberspace itself. NATO has the nascent elements of such a bifurcated approach in place with its Allied Command Operations for the immediate issues and Allied Command Transformation for the future issues related to conflict in cyberspace. Without such a comprehensive view, planners and decision makers add risk in their activities by not characterizing the full spectrum of the interactions and effects of creation and operation in cyberspace.

REFERENCES

[1] "Cyberspace Policy Review: Assuring a Trusted and Resilient Information and Communications Infrastructure." Washington, DC: The White House, May 2009.

[2] T. L. Thomas. "Google Confronts China's 'Three Warfares'." *Parameters*, vol. 40, no. 2, pp. 101-113, Summer 2010.

[3] S. W. Korns and J. E. Kastenberg. "Georgia's Cyber Left Hook." *Parameters*, vol. 38, no. 4, pp. 60-76, Winter 2008-2009.

[4] W. F. Lynn III. "Defending a New Domain: The Pentagon's Cyberstrategy." *Foreign Affairs*, vol. 89, no.5, pp. 97-108, Sep./Oct. 2010.

[5] "U.S. Cyber Command Fact Sheet." Fort Meade, MD: U.S. Cyber Command Public Affairs, Oct., 2011.

[6] "Department of Defense Strategy for Operating in Cyberspace." Washington, DC: Dept. of Defense, Jul. 14, 2011.

[7] "Defending the Networks: The NATO Policy on Cyber Defence." Brussels, Belgium: NATO Public Diplomacy Division, Jun. 2011.
[8] S.M. Bellovin et al. "Can It Really Work ? Problems with Extending EINSTEIN 3 to Critical Infrastructure." *Harvard National Security Journal*, vol. 3, pp.1-38, 2011.
[9] A. M. Denmark and J. Mulvenon. "Contested Commons: The Future of American Power in a Multipolar World." Washington, DC: Center for a New American Security, Jan. 2010.
[10] D. T. Kuehl. "From Cyberspace to Cyberpower: Defining the Problem." in *Cyberpower and National Security*, Washington, DC: National Defense University Press and Potomac Books, 2009, pp. 24-42.
[11] J. L. Caton. "Cyberspace and Cyberspace Operations." in *Information Operations Primer*, AY12 ed., Carlisle, PA: U.S. Army War College, Nov. 2011, pp. 19-32.
[12] "Enabling Distributed Security in Cyberspace: Building a Healthy and Resilient Cyber Ecosystem with Automated Collective Action." Washington, DC: Depart. of Homeland Security, Mar. 23, 2011.
[13] A.T. Mahan. *The Influence of Sea Power Upon History 1660-1783*. Mineola, NY: Dover, 1987 reprint, pp. 30, 31-32.
[14] S. M. Barney. "Innocent Packets? Applying Navigational Regimes from the Law of the Sea Convention by Analogy to the Realm of Cyberspace." *Naval Law Review*, vol. 48, pp. 58-87, 2001.
[15] "2010 Report to Congress of the U.S.-China Economic and Security Review Commission." Washington, DC: U.S. Government Printing Office, Nov. 2010, pp. 243-244.
[16] "The 'Wild and Woolly' World of Bulletin Boards." *All Things Considered*, National Public Radio, Nov. 21, 2009.
[17] K. Koscher *et al*. "Experimental Security Analysis of a Modern Automobile." presented at the 2010 IEEE Symp. On Security and Privacy, Oakland, CA, May 2011.
[18] "Finding Regarding the Market Events of May 6, 2010: Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues." Washington, DC: U.S. Commodity Futures Trading Commission and U.S. Securities and Exchange Commission, Sep. 30, 2010.
[19] Johnson et al. "Financial black swans driven by ultrafast machine ecology." technical working paper, Cornell University Library, Ithaca, NY, Feb. 2012.
[20] E. Tyugu. "Artificial Intelligence in Cyber Defense." *Proc. 2011 3rd Conf. on Cyber Conflict*, NATO Cooperative Cyber Defence Centre of Excellence, Tallinn, Estonia, 2009.
[21] "Active Engagement, Modern Defence: Strategic Concept For the Defence and Security of The Members of the North Atlantic Treaty Organisation." adopted by Heads of State and Government in Lisbon, Portugal: NATO, Nov. 19, 2010.
[22] B. Obama. "International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World." Washington, DC: The White House, May 2011.
[23] J. B. Michel et al. "Measured Responses to Cyber Attacks Using Schmitt Analysis: A Case Study of Attack Scenarios for a Software-Intensive System," *Proc. of Twenty-seventh Annu. Int. Software and Applications Conf.*, Dallas, TX: IEEE, Nov., 2003.
[24] D. Bilar. "On nth Order Attacks." *Proc. 2009 Conf. on Cyber Warfare*, NATO Cooperative Cyber Defence Centre of Excellence, Tallinn, Estonia, 2009.
[25] G. Klein *et al*. "Enhancing Graph-based Automated DoS Attack Response." *Proc. 2009 Conf. on Cyber Warfare*, NATO Cooperative Cyber Defence Centre of Excellence, Tallinn, Estonia, 2009.
[26] L. Beaudoin, N. Japkowicz, and S. Matwin. "Autonomic Computer Network Defence Using Risk State and Reinforcement Learning." *Proc. 2009 Conf. on Cyber Warfare*, NATO Cooperative Cyber Defence Centre of Excellence, Tallinn, Estonia, 2009.
[27] E. Skoudis. "Information Technology and the Biotech Revolution." in *Cyberpower and National Security*, Washington, DC: National Defense University Press and Potomac Books, 2009, pp. 241-250.
[28] S. H. Starr. "Toward a Preliminary Theory of Cyberpower." in *Cyberpower and National Security*, Washington, DC: National Defense University Press and Potomac Books, 2009, pp. 43-88.
[29] N. J. DeLong-Bas. "The New Social Media and the Arab Spring." *Oxford Islamic Studies Online*, Jun. 2011.
[30] O. Thonnard, W. Mees, and M. Dacier. "Behavioral Analysis of Zombie Armies." *Proc. 2009 Conf. on Cyber Warfare*, NATO Cooperative Cyber Defence Centre of Excellence, Tallinn, Estonia, 2009.
[31] D. Sornetter. "Dragon-Kings, Black Swans and the Prediction of Crises." *Int. J. of Terraspace Sci. and Eng*., pp. 1-18, 2009.
[32] R. Temmingh and K. Geers. "Virtual Plots, Real Revolution." *Proc. 2009 Conf. on Cyber Warfare*, NATO Cooperative Cyber Defence Centre of Excellence, Tallinn, Estonia, 2009.
[33] C. Koch and K. Hepp. "The relation between quantum mechanics and higher brain functions: Lessons from quantum computation and neurobiology." California Institute of Technology, Pasadena, CA, Apr. 2007.
[34] R. Kurzweil. *The Singularity is Near: When Humans Transcend Biology*. New York, NY: Viking, 2005.
[35] P. Teilhard de Chardin. *The Phenomenon of Man*. English translation by Bernard Wall, New York, NY: Harper Brothers Publishers, 1959, pp. 18, 182.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Applying Traditional Military Principles to Cyber Warfare

Samuel Liles
Cyber Integration and Information Operations Department
National Defense University iCollege
Washington, DC
Samuel.Liles@NDU.edu

Marcus Rogers
Computer and Information Technology Department
Purdue University
West Lafayette, IN
rogersmk@purdue.edu

J. Eric Dietz
Purdue Homeland Security Institute
Purdue University
West Lafayette, IN
jedietz@purdue.edu

Dean Larson
Larson Performance Engineering
Munster, IN
deanlarson@larsonperformance.com

Abstract: Utilizing a variety of resources, the conventions of land warfare will be analyzed for their cyber impact by using the principles designated by the United States Army. The analysis will discuss in detail the factors impacting security of the network enterprise for command and control, the information conduits found in the technological enterprise, and the effects upon the adversary and combatant commander.

Keywords: *cyber warfare, military principles, combatant controls, mechanisms, strategy*

1. INTRODUCTION

Adams informs us that rapid changes due to technology have increasingly effected the affairs of the military. This effect whether economic, political, or otherwise has sometimes been extreme. Technology has also made substantial impacts on the prosecution of war. Adams also informs us that information technology is one of the primary change agents in the military of today and likely of the future [1]. There is a difference between using information technology or cyber space as a domain to fight and fighting in the domain of cyber space. Some of the differences appear to be maturity issues in understanding the cyber space domain. The translation of warfare strategies from other domains into an operational art is a process that is simply in its infancy [2]. General Alexander in 2007 said that we currently face many similar issues grappling with cyberspace as a war-fighting domain as the military did during the Interwar years from 1919 to 1938 understanding air-power [2].

This lack of maturity in understanding cyber space appears to be related to other myths of conflict. There are four myths of future land war suggested by Dunlap that are easily applied to cyber warfare; 1) Our most likely future adversaries will be like us; 2) We can safely downsize our military in favor of smaller, highly trained forces equipped with high-technology weapons; 3) We can achieve information superiority and even dominance in future conflicts; 4) Modern technology will make future war more humane if not bloodless. These myths are based the larger quandary known as the "revolution in military affairs" and the "generational constructs" being developed at the same time as it was written by Dunlap [3].

Cyber warfare has many definitions which makes it hard to state exactly what it is when it is many things depending on point of view. One suggested definition is that cyberwar is conducting military operations according to information-related principles while disrupting, destroying and knowing much about an adversary while keeping them from knowing about you [4,5]. Land warfare though has a very similar definition, as we will see in much deeper detail later. This leads into the purpose and scope of this paper:

> Using the conventions for land warfare, what kinds of cyber threats constitute attacks and how do we characterize possible cyber warfare scenarios or attack techniques to provide concepts for a generalized approach that supports situational awareness of the cyber battle space or "terrain"? How does this tool vary for first responders or military operations?

As such it might help to discuss the basic principles of the preeminent land war force in the world. It helps to understand the scope if the principles are detailed. The United States Army in dealing with land warfare has nine principles of war:

> 1) Objective – direct every military operation towards a clearly define, decisive, and attainable objective; 2) Offensive – seize, retain, and exploit the initiative; 3) Mass – concentrate the effects of combat power at the decisive place and time; 4) Economy of force – allocate minimum essential combat power to secondary efforts; 5) Maneuver – place the enemy in a disadvantageous position through the flexible application of combat power; 6) Unity of command – for every objective, ensure unity of effort under one responsible commander; 7) Security – never permit the enemy to acquire an unexpected advantage; 8) Surprise – strike the enemy at a time or place or in a manner for which he is unprepared; 9) Simplicity – prepare clear, uncomplicated plans and clear, concise orders to ensure thorough understanding. [6]

If mass and economy are related it is important for the combatant commander to understand how cyber enables the mission. The network centric aspects of future battle spaces means that a new weakness has been included too. Effective employment of cyber assets includes an understanding of defending those assets [7]. Parks details several principles of cyber warfare including that cyber warfare must have kinetic effects [8]. Discussing this, Parks says, there are no laws in cyber space, somebody can do just about anything to somebody else given enough authority, tools are dual use, defender and attackers control very little, and cyber space is not consistent. Parks illustrates some of the differences between what the Army doctrine would expect and the capabilities of actual cyber space. Saydjari also looked at the corresponding relationship between information assurance and military doctrinal statements

[9]. Saydjari states that cyber warfare relies on: sensors and exploitation; situational awareness; defensive mechanisms; command and control; strategies and tactics; and then finally science and engineering. The question of effectiveness of attack is in doubt when there is a substantial disconnect between published Army doctrine and the experts opinions on how it all fits together. Attacks from cyber space are cheaper and have substantial impediments to attribution, and as such it is not hard to believe that adversaries of a nation state could attack using information technology in an attempt to manipulate policy and decision makers [10]. Brooks suggested that information operations as a discipline needed to be included in the primary planning phases of operations. Information operations are a form of attack that still fits within the nine principles of military doctrine [10]. This is exactly what China was accused of doing on numerous occasions. Though it appears in most cases infiltration of networks by technology or human agents is done for the exfiltration of information (espionage) [11-15]. Of course, there is also the threat of other nation states such as Russia engaging in espionage through the network [16]. This is not to say that the United States is not also involved in espionage activities. Corn explains that the Pentagon has examined computer communications in transit to determine the modes of operations and goals of fringe groups [17].

The forms of attack are varied and inclusive of goals other than simply winning territory. Conflict is a continuum of strategies into which insurgency rises as a primary strategy. As such irregular warfare and insurgency are old ideas that get applied to new domains of battle repeatedly [18]. The distinctions between irregular warfare, insurgency, low-intensity conflict, guerilla warfare, and terrorism are counterpointed by the merits of each on a continuum of conflict. Gray reminds us that war is basically and simplistically war. The rules of war are applied often after the conflict [18].

Asymmetry, the defining element of insurgency, is not designed to win in the battle-space but to disrupt, distract, disconnect, or debilitate the nation state [19-21]. Relatively speaking the global communications network is nearly exclusively an asymmetric environment where mass and maneuver have minimal meaning. Dion examines the impact of digital capabilities in bringing mass and maneuver to the battle space [19]. This though is a capability not a weapon. Dion is discussing the layering of the digital information technology environment upon the weapons platforms of the Army. This gives the nation-state a significant information edge over the adversary. Layering cyber space capabilities onto terrestrial weapons platforms is not functionally different from using naval forces to support land forces. Another example might be space assets, such as reconnaissance satellites, that support all natural domains (air, land, sea) similar to how cyber supports command and control.

Tying back to the tenets espoused previously, Groh sees military conduct in cyber space as network centric operations and reflecting back to the original tenets of Army doctrine [22]. Specifically he has four information centric statements paraphrased as: 1) Robust networked force improves information sharing; 2) Information sharing and collaboration enhance the quality of information and situational awareness; 3) Shared situational awareness enables self-synchronization; 4) These all increase mission effectiveness. Each point can be brought back to the ideas of speed, maneuver, and unity of command. In this regard network centric warfare is specifically linked to these concepts. As such cyber warfare, which is attacking those channels of information flow, will target the nodes of communication. If taken as information operations

centric, there is some worry of overstating the case. Groh specifically warns that network centric warfare is not a silver bullet as his tenets of network centric warfare limit the doctrinal application to a few areas of specialty.

2. SITUATIONAL AWARENESS TO INFORMATION AS CONTESTED TERRAIN

Cyberspace is not a wholly new area of conflict and is not necessarily a new or nonphysical construct. In fact it is a wholly physical construct much like any other terrain [23]. The advent of cyberspace as a contested domain has significant implications to military doctrine. The strategic understanding of impacts, such as situational awareness removing the fog of war from commanders' current understanding of conditions, are nearly incomprehensible. The strategic and cognitive impacts to leaders' planning and operational capability should be extensive [24,25].

Command and control warfare is the application of computer information technology for offensive and defensive military operations. Rather than being a primary mode of operations, command and control warfare is an enhancement to the ability of the military unit to operate [26,27]. The cyber assets used by a commander to control can also be used against the commander. As such there is an inherent linkage between the communication infrastructure and the combatant commander. Though there is a relative desire on the part of technologists to say computer information technology it might be important to note that information technology and computers exist at all levels and not simply the desktop personal computer. Many military radios and encryption systems are filled with computers too.

The addition of information technology and computerized capability incurs a set of new risks that are balanced alongside the gains of the new technology. Critics of the technology may overstate the risks. One element likely overstated is the preponderance of "collapse theory" as the primary risk associated with increased information technology capability [24]. Large scale computing systems and communications systems are built with redundancy and scalable capacity. Overwhelming these systems is possible but the idea of collapse theory is that they will not recover from failure.

The ability to utilize ubiquitous computing for decision support and communication through the battle space has substantially increased the scope and vision of the commander in what is becoming known as network centric warfare [28]. There are five tenets to the process of waging network centric warfare according to Adkins 1) Knowledge of the competition, or in the case of the military, the adversary; 2) Near real time shared situation awareness; 3) Communications of the corporate or commander's intent; 4) Decentralized execution of plans; 5) Enabling self-synchronization [28]. This is expanding once again the capability from simply information operations (attacking information flows), past command and control warfare (attacking commanders intent), to utilizing the network to enhance the commander's control. Usually though we see command and control warfare as a strategy to disrupt decision processes.

Command and control in warfare is a strategic issue and tactical conundrum as network

centric capability is realized, though, it is not fully realized, or equally realized across the military enterprise. Acquisition of capability that was commercially available but not within the procurement system slowed and degraded the capability of the Army in Operation Iraqi Freedom. This created an expectation gap of possible versus the operational [29]. Examining this issue in depth Cogan also detailed that tactical communications were degraded by the capability of the end point equipment versus the capability of the backbones bandwidth. From this examination we can deduce two clues about attacking command and control from a cyber warfare denial of service aspect. First, the war, even with degraded capability of the networked equipment, was waged rapidly and successfully. Second, the acquisition process had more effect on the Army capability than the meager attempts to destroy or infiltrate the network. This would be counter to the theorists of collapse theory as discussed by Leonhard [30].

This has left the command and control aspects of warfare much where they were two or three decades ago. Rather than a decrease in capability, the expectations simply have not been met. Where there is increased capability it is held up as an example of superiority. If command is carried out by direction, by plan, or by influence has the automated nature of command and control met those tenets [31]? Command by direction being the oldest method of command, and command by influence being a relatively new construct suggests some maturation of the process. Into this mix cyber warfare as a capability is added.

Metaphors of attack often lack realistic operational thinking. The colloquialism that all elegant metaphors degrade under enough pressure surely must hold true. A favored metaphor of layered defense, or defense in depth, may make metaphorical sense but can be problematic in reality. This is an issue between the logical structure of networks and the physical structure of them. A castle metaphor is good to discuss computer and network security but it lacks certain elegance and sophistication of thinking. Empirical research suggests that layered defense strategies consistently decrease the security of a system. This is based on the increased complexity and increased control services that an adversary could attack [32]. So not only do the cognitive issues degrade but the actual security mechanisms may be degrading, too.

There is also the logical layer in how technology is used. Information systems exist to allow people to communicate and coordinate activities much like any form of technology based communication. Information technology though has some issues with how communication is conducted. Social media and information systems can be exploited through the systems' inherent human centric lag [33]. As an example an insurgency is an inherently social organization with a political purpose. As such a social network approach to understanding them can give clues as to how they are using technology and what that interaction might look like in the real world. Insurgencies are a particular subset of the spectrum of conflict and defy rigid classification [34]. So, the logical and cognitive layers may be both supported by information technology and then exploited (used) by adversaries alike.

One of the issues to the Army and other military organizations is the simple prevalence of the technologies necessary to wage war in cyber space. This is a social problem using technology and not a technology enabling social interaction [35]. The technology in some cases has become the reason rather than the use of the technology. In other cases technology is banned because it is technology rather than the behavior of the misuse. This conundrum has opened

up avenues to exploitation not previously exposed. Whether considered from the prospect of actually using cyber space as a tool to attack, or more likely using cyber space tools to coordinate and communicate a highly desirable capability exists. The ability to raise a mass of socially, technically, networked people with defined purpose is the new *Levee en Masse* [17]. Unfortunately large organizations rarely have the ability to leverage this capability as fast as smaller organizations.

3. APPROACHES TO AN ATTACK IN CYBER SPACE

There are specific behaviors and paths that an attacker will usually take. "An attacker is going to attempt to deny, corrupt, or exploit the adversary's information or influence the adversary's perception" [20]. There is a pretty standard process that will accomplish the prior. The adversary will gather information about the target, plan the attack, and execute the attack. This process is similar to any military activity and only the depth of each step and the conclusions might be different between traditional arms and cyber attacks. Currently there is little in the way of a cyber war rules of engagement. Related to this gap is the missing legal and doctrine development for waging cyber warfare by nation states [36]. The process can take into account each of the nine principles and may be tightly organized around a cross domain approach (utilizing tactics from multiple avenues of attack not simply cyber). This leads to a discussion on strategy and what it means to those nine principles.

For the purpose of considering strategic information warfare Rattray describes three forms of attack: 1) mechanical attacks; 2) electromagnetic attacks; 3) digital attacks. Each of these forms of attack takes on specific strategic aspects and merits [23]. Each of the forms of attack can be directed at or from cyber as the operating weapons system. When considering the merits of attack and defense in the cyber battle space the normal frictions of combat become elusive. Most military doctrine currently understood is about war of attrition, but cyber warfare does not seem as weak to cessation of communication as previously thought [37]. Working around technical disruptions has continued without much in the way of the issue moving forward as a prelude or cyber attack. Various systems and methods of design and infrastructure have been examined to determine an appropriate strategy for dealing with outages [38]. So, even if the attack is successful it may be seen as degradation before it is seen as a serious issue.

When the combatant commander contemplates attack there are serious issues to consider. There is a caution to combatant commanders during the attack phase of command and control warfare to steer clear of imitative deception to commit perfidious acts (false flag operations) as these could be considered war crimes [39]. How this may actually be built into battle plans is not currently discussed outside of classified environments. Actually, not much is discussed in unclassified environments about military training in cyber space. The training of military computer attack teams are classified, but due to the open nature of the technologies involved are likely similar to any other corporate red team capability [40].

4. GENERATIONAL CONSTRUCTS ATTEMPT TO DEFINE CYBER CONFLICT

The revolution in military affairs in many ways is the root of the substantial change and advancement of generational constructs to explain war theory since the mid 1990s [41]. One of the newer concepts suggested is the idea of generational constructs to define conflict strategies and capabilities. Each of the generations of warfare is defined as a capability, technology, or tactic that builds upon the previous generation. For this paper a detailed discussion is not within the scope but see other works by the author for that examination. The concepts and movement of ideas about generational constructs continues to today with work by Hammes. Hammes expands his concepts of generational constructs from fourth to a possible fifth generational component. This fifth generational component is an information operations and cyber enabled population's conflict realm [42]. This work is in addition to the work he did in 2004 where fourth generational related insurgency specific constructs were detailed and analyzed.

Hammes discusses in depth the changing face of war and details the generational warfare construct as an explanatory mechanism. Rather than thinking temporal, each succeeding generation of warfare is advancement in methodology. The first two generations of warfare are answers to technical problems with technology solutions [43]. The third generation of warfare is a change in tactics as Hammes suggests evidenced by mechanization and speed of armor allowed to flourish during World War 2 during the German invasion of Poland[43]. For our purposes in considering the addition of cyber conflict the fourth generation as population centric is especially of interest. The conflict space of fourth generation warfare is that of insurgency or populist aggression against the nation states as Hammes illustrates while discussing Mao [43]. Hammes (2007) builds upon the former to add a cyber and information spectrum for a fifth generational construct.

The realm of cyberspace allows for the fourth generation warfare construct to grow rapidly. When considering the Maoist "displacement strategy" of building "parallel hierarchies" government legitimacy is threatened [44]. Rather than relying on the traditional elements of military warfare such as maneuver, the insurgent in cyber space can use temporal displacement to negate nation state power. The nation-state though should be especially careful as the technological advantage can be lost in a societal shift [45]. Terrorism is especially linked to the idea of legitimacy. Thinking back to the previous discussion on asymmetry when mass and maneuver or not a capability the adversary can leap past them to take on legitimacy of governance. Terrorism via cyber means may break the principles back.

Cyber terrorism as discussed is a relatively inexpensive tool to use in an attack. Yet is wholly an expensive and difficult activity to protect against. Though skeptical Giacomello discussed cyber terrorism in detail as a possibility rather than defined capability [46]. One issue detailed by Giacommello is that the word terrorism is relatively meaningless being defined differently in law and literature. This is supported by in depth by Gordon [47]. In considering the merits of cyber terrorism Giacommello makes a startlingly conclusion that the issue is primarily a cultural phenomenon rather than technical. Perhaps not nearly as startling as expected, as all conflict regardless of the tools is likely cultural in nature.

5. MILITARY OPERATIONS IN CYBER SPACE

Discussing the issues of information in the battle space is nothing new to the Army [45]. There has, however, been a growing scholarship of dealing with information operations from the standpoint of conflict communications. There is also prevalent thread of thought in the international community that suggests information operations can decrease the perfidy of conflict [48].

Simply having computers and using them as communication conduits is not the only issue to combatant commanders considering cyber conflict. The ethics and assumptions of actions taken in cyber space especially computer network attack must be considered. A combatant commander must consider the ideas of discrimination between targets and proportionality of response. [39,49]

6. RESULTS

Coming back to the discussion of how the Army defines conflict and the nine principles of war and combat discussed previously, a series of resulting conclusions can be mapped. These are by no means expected to be the only conclusions that could be derived from the literature. They however do map and can be seen through the lens of the literature. As a cyber conflict space these nine principles have specific allegory to the cyber domain.

The objective in cyber conflict has not substantially changed from the previous consideration of terrestrial conflict. The idea of what attack means and the means of that attack has not substantially changed. The use of generational constructs and information operations has not substantially changed the concept of defining a goal or end-state to an engagement. Relatively simple in statement the where withal to accomplish the task through cyber means can be harder to determine. One element to objective that should not be ignored is that the set of strategic targets and objectives with cyber has been substantially increased in scope.

Taking the offensive is an interesting question. In the idea of generational warfare constructs and low-intensity-conflict, which is related to the tactical choice of insurgency, the offensive may not be similar to previous engagements. To be more specific the forms of conflict are likely to relate more to the fourth and fifth generational models suggesting insurgency and less to high-intensity conflict models where other principles relate closer. It appears taking the offensive may itself be in doubt as limiting war to cyber space may make the principle of offense less obvious. The roles of offense and defense seem to blur within an insurgency model as they do within cyber space.

Mass and economy of force as stated earlier appear to be related within the literature when considering the significant asymmetry of attack strategies and defense requirements. As such, examples of mass jump to the forefront that may not be the best examples. A distributed denial of service appears to be mass when in actually the result is significant but the force behind it is not. That might suggest that technology itself is a force multiplier and in the case of computer information technology substantial. However, that also misses the point that the

effect is primarily against other computer information technology. The user of armor or heavy weapons is technology that has significant impact against people. To be a relevant principle in considering mass and economy of force they would have to effect people. Unfortunately to gain that effect a third element must be rolled in and mass subtracted from the equation. As principles to broadly define military strategic issues they are weakening quickly.

A principle of maneuver exists, as it is not a fact of physical, but also emotional and cognitive. The previously discussed information operations use maneuver to speed combatant commanders and adversaries decisions cycles into appropriate resulting conclusions. Already defined for us through the information operations aspect of current strategic thought we can now apply that same principle of maneuver even faster through computer information technology.

Unity of command as a principle we saw from the literature strongly holds to the use of information technology as the current tool suite used. As seen in several cases command and control are inherently part of this equation and acted upon by computer information technology assets. Those assets are inherently part of the current landscape and the concepts of network centric warfare within the literature are deeply rooted to this basic principle. It then follows that unity of command is a fundamental principle of cyber warfare as it is currently used within computer information technology. Unity of command has used technology for the idea of command and control since smoke signals, semaphore and watch towers as beacons. The advent of computer information technology has only made the cyber landscape faster.

Without a lack of security the computer information technology attack vector might be said to be missing. Unfortunately perfectly perfected computing systems are still perfectly exploitable by people using them for purposes exactly as designed with nefarious results. The literature describes in detail the ideas of cascading failures and the criticisms of that flawed logic. What are not described are insider actions by military entities such as spies and agents. That is likely a classified discussion but a relevant thread for future research.

The act of surprise grows harder and more difficult on the high intensity conflict terrains of the modern battlefield. Observance of the last several incursions by foreign and domestic powers into other sovereign territory have been preceded by massive buildups where the actual attack appears as a pressure cooker finally blowing off steam. Surprise might be characterized as, that it took so long, instead of actually being stealthy. In the computer information technology domain of cyber warfare it becomes rapidly obvious that many attacks are taking place daily. This is supported by numerous literature resources that described earlier the idea of security being lacking. Thus surprise has much to be compared to current terrestrial combat.

Simplicity is in the binary. There is little simpler than the binary of on-off that runs computers. Refuting that point is the systems of systems discussion identified in the literature, which suggested massive scalable systems are created with significant holes in their security. The literature would support that the simplicity assists the adversary through the other principle of economy of force, and that the attacker garners the benefit while the defender is on the opposite side of the simplicity coin. The principle of simplicity as identified in the literature though cascading systems failure and systems of systems approach to design must support the attacker more than it will the defender.

7. CONCLUSIONS

Looking at the conventions of land warfare and the principles of war that constitute strategy and tactics it becomes obvious that there is a substantial disconnect when considering cyber warfare. In fact, there are those who simply say it does not exist [39]. A disconnect between the legal, moral, and ethical considerations perhaps: the conventions for land warfare often refer to the laws of land war, as in the Geneva Convention. However in answering the research question, the author decided to focus primarily on the second part of the research question to answer how the techniques and concepts for generalized approaches to situational awareness might be accomplished.

In ignoring the first part of what constitutes an attack under the law of war, we were able to talk about a variety of attacks. The discussion within this paper answers the idea of attack centered on the types of attack that were possible. Part of this is that perfidy and *jus in bello* in information security simply has not been described succinctly [39]. Simply put the use of the civilian network which is nearly a requirement puts the entire first part of the original research question into a quandary. The civilian network component as described adds possible perfidy to every attack and a nearly defacto risk of violations of the laws of war [25,50].

Finally the last part of the question of how this tool varies is easily answered as discussed previously. The attack is always going to be at an asymmetric advantage that cannot be substantially changed. The level of effort to enter the field of battle no longer requires the nation state. As such the first responder is radically empowered by the scope of their capability to attack but have no real capability at defense when integrated into a corporate or military information enterprise. This is the asymmetric advantage that currently does not erode or seem to erode under scalable systems.

As such the research question has been answered in detail with supporting literature from a variety of resources.

REFERENCES:

[1] T. K. Adams, “Radical destabilizing effects of new technologies,” Parameters, vol. 1998, pp. 99-111, 1998.
[2] K. B. Alexander, “Warfighting in cyberspace,” Joint Forces Quarterly, vol. 3rd Quarter, pp. 58-61, 2007.
[3] C. Dunlap, “21st century land warfare: Four dangerous myths,” Parameters, vol. 1997, pp. 27-37, 1997.
[4] J. Arquilla and D. Ronfeldt, Networks and netwars: The future of terror, crime, and militancy. Santa Monica, CA: RAND, 2001.
[5] B. Panda and J. Giordano, “Defensive information warfare,” Communications of the ACM, vol. 42, pp. 31-32, July 1999.
[6] U. S. Army, “FM 3.0 Operations,” T. U. S. Army, Ed., ed. Washington DC, 2001, p. 104.
[7] P. Murdock, “Principles of war on the network-centric battlefield: Mass and economy of force,” Parameters, vol. 2002, pp. 86-95, 2002.
[8] R. C. Parks and D. P. Duggan, “Principles of cyber-warfare,” in 2001 IEEE Workshop on Information Assurance and Security, United States Military Academy, West Point, NY, 2001, pp. 122-125.
[9] S. Saydjari, “Cyber defense: Art to science,” Communications of the ACM, vol. 47, pp. 53-57, March 2004.
[10] P. Brooks, “A vision of PSYOP in the information age,” Special Warfare, 2000.

[11] S. Cooper. (2006) China's secret war. Popular Mechanics. Available: http://www.popularmechanics.com/technology/military_law/3319656.html
[12] J. A. Lewis, Computer espionage, Titan Rain, and China. Washington DC: Center for Strategic & International Studies, 2005.
[13] T. Espiner. (2005, November 17, 2007). Security experts lift lid on Chinese hack attacks. Available: http://news.zdnet.com/2100-1009_22-5969516.html
[14] T. Luard. (2005, November 16). China's spies come out from the cold (International Version ed.). Available: http://news.bbc.co.uk/2/hi/asia-pacific/4704691.stm
[15] D. Sevestopulo. (2007, November 17). Chinese military hacked into Pentagon. Available: http://www.ft.com/cms/s/0/9dba9ba2-5a3b-11dc-9bcd-0000779fd2ac.html?nclick_check=1
[16] B. Drogin, "Russians seem to be hacking into Pentagon: Sensitive information taken--but nothing top secret," in SFGate.com, ed. San Francisco, CA, 1999.
[17] D. Corn. (1996) Pentagon trolls the net. The Nation.
[18] C. S. Gray, "Irregular warfare: One nature, many characters," Strategic Studies Quarterly, pp. 35-57, 2007.
[19] E. Dion, "The e-Forces!: The evolution of battle-groupings in the face of 21st century challenges," Canadian Army Journal, p. 3, October 29-30 2004.
[20] A. J. Elbirt, "Information warfare: Are you at risk," IEEE Technology and Society Magazine, pp. 13-19, 2003.
[21] T. Franz, M. Durkin, P. Williams, R. Baines, and R. Mills, "Defining information operations forces," Air & Space Power Journal, pp. 1-11, 2007.
[22] J. L. Groh, "Network-centric warfare: Leveraging the power of information," in U.S. Army War College Guide to National Security Issues. Third Edition. vol. 1, ed Carlisle, PA: Army War College: Strategic Studies Institute, 2008, pp. 323-338.
[23] G. J. Rattray, Stategic warfare in cyberspace. Cambridge, Massachusetts: MIT Press, 2001.
[24] R. R. Leonhard, "A culture of velocity," in Digital war: A view from the front lines, R. L. Bateman, Ed., ed Novato, CA: Presidio Press, 1999, pp. 131-152.
[25] R. C. Molander, A. S. Riddile, and P. A. Wilson, "Strategic information warfare: A new face of war," Parameters, vol. 1996, pp. 81-92, 1996.
[26] R. F. Erbacher, "Extending command and control infrastructures to cyber warfare assets," in Workshop on Information Assurance and Security United States Military Academy, West Point, NY, 2005.
[27] S. Paradis, A. Benaskeur, M. Oxenham, and P. Cutler, "Threat evaluation and weapons allocation in network-centric warfare," in 7th International Conference on Information Fusion (FUSION), 2005, pp. 1078-1085.
[28] M. Adkins, J. Kruse, and R. Younger, "Ubiquitous computing: Omnipresent technology in support of network centric warfare," in 35th Hawaii International Conference of Systems Sciences, Hawaii, 2002, p. 9.
[29] K. J. Cogan, "A view of command, control, communications, and computer architectures at the dawn of network centric warfare," Issue Paper Center for Strategic Leadership, vol. 2-07, 2007.
[30] R. R. Leonhard, The principles of war for the information age. New York: Presidio Press, 1998.
[31] T. J. Czerwinski, "Command and control at the crossroads," Parameters, vol. 1996, pp. 121-132, 1996.
[32] D. L. Kewley and J. Lowry, "Observations on the effects of defense in depth on adversary behavior in cyber warfare," in Workshop on Information Assurance and Security, United States Military Academy, West Point, NY, 2001.
[33] A. Fuxman, P. Giorgini, M. Kolp, and J. Mylopoulos, "Information systems as social structures," in The International Conference on Formal Ontology in Information Systems, Ogunquit, Maine, USA, 2001, pp. 3-9.
[34] B. Reed, "A social network approach to understanding an insurgency," Parameters, vol. 2007, pp. 19-30, 2007.
[35] N. Schachtman. (2007, October 1). How technology almost lost the war: In Iraq the critical networks are social - not electronic. Available: http://www.wired.com/politics/security/magazine/15-12/ff_futurewar
[36] D. B. Hollis, "Title," unpublished|.
[37] M. Herman, "Modeling the revolution in military affairs," Joint Forces Quarterly, vol. Autumn/Winter, pp. 85-90, 1998-99.
[38] B. Hicks, "Transforming avionics architecture to support network centric warfare," in The 23rd Digital Avionics Systems Conference, 2004.
[39] W. J. Bayles, "The ethics of computer network attack," Parameters, vol. 2001, pp. 44-58, 2001.
[40] J. Lasker. (2005, October 1). U.S. military's elite hacker crew. Available: http://www.wired.com/politics/security/news/2005/04/67223
[41] R. Bunker, "Generations, waves, epochs: Modes of warfare and the RPMA," Airpower Journal, pp. 1-10, 1996.

[42] T. X. Hammes, “Fourth generation warfare evolves, fifth emerges,” Military Review, vol. May-June, pp. 14-23, 2007.
[43] T. X. Hammes, The sling and the stone: On war in the 21st century. St. Paul, MN: Zenith Press, 2004.
[44] D. Kilcullen, The accidental guerrilla: Fighting small wars in the midst of big ones. Oxford: Oxford University Press, 2009.
[45] R. J. Harknet, “Information warfare and deterrence,” Parameters, vol. 1996, pp. 93-107, 1996.
[46] G. Giacomello, “Bangs for the buck: A cost benefit analysis of cyberterrorism,” Studies in conflict & terrorism, vol. 27, pp. 387-408, 2004.
[47] S. Gordon, “Cyberterrorism?,” ed. Cupertino, CA: Symantec Corporation, 2003, p. 15.
[48] L. Armistead, Information operations: Warfare and the hard reality of soft power. Washington, DC: Brassey’s, Inc., 2004.
[49] E. V. Leighninger, “Is software warfare d'unthinkableĒ? or is there a rational basis for its adoption?: A proposal for ethical reflection and action,” ACM SIGSAC Review, vol. 9, p. 28, 1991.
[50] J. Mulvenon, “Toward a cyberconflict studies research agenda,” IEEE Security and Privacy, pp. 52-55, 2005.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

The Principle of Maneuver in Cyber Operations

Scott D. Applegate
Volgenau School of Information Technology
Engineering
George Mason University
Fairfax, Virginia
sapplega@gmu.edu

Abstract: The United States Military describes the concept of maneuver as the disposition of forces to conduct operations by securing positional advantages before and or during combat operations. This paper will briefly explore how the concept of maneuver in kinetic operations has evolved over time and how that concept relates to cyber operations and cyber warfare. It will attempt to define what constitutes the principle of maneuver within cyberspace as it relates to the traditional concept of maneuver in warfare and how the unique domain of cyberspace alters this concept. This paper will explore the characteristics of maneuver in cyberspace and the basic offensive and defensive forms of maneuver that have thus far emerged will be identified and analysed. The author will also briefly touch on the issue of sovereignty in cyberspace as it relates to cyber maneuver and attempt to identify how and when the concept of cyber maneuver might cross the line to violate a state's sovereignty. This paper will demonstrate that there is a valid concept of maneuver in cyberspace, and that the stealth and anonymity provided by the Internet allows for blatant acts which, in a kinetic operation, would most like result in open armed conflict.

Keywords: *cyber conflict, cyber maneuver, cyber operations, cyber warfare*

1. INTRODUCTION

Military Strategists have been writing on the principles and characteristics of warfare for more than two thousand years. Although the specific principles differ over time and in relation to particular strategists, the principle of maneuver has been an important concept and has been a determining factor in warfare since some of the earliest recorded battles. As technology has evolved and allowed for the expansion of warfare into new domains, so too has the concept of maneuver changed. The exploration of the seas created a new unique domain and introduced the concept of a global commons, bringing with it new challenges to overcome. Air and Space added a new dimension to the principle of maneuver and caused yet another shift in military strategy. During the last two decades, the introduction of computing systems and the Internet formed an interconnected, virtual environment that has led to the designation of a fifth warfighting domain

known as Cyberspace. This new domain has its own set of unique characteristics and challenges and significantly overlaps operations in all four of the other warfighting domains.

The United States Military describes the concept of maneuver as the disposition of forces to conduct operations by securing positional advantages before and or during combat operations [1]. While this description has some applicability to operations in the domain of cyberspace, it is clear that this open, borderless, virtualized environment alters this principle significantly and an effort must be undertaken to understand and codify these changes.

While cyberspace is considered a warfighting domain, thus far it has manifested itself as more of a contested domain characterized by constant conflict between various competitor states, non-state actors and private entities. Battles rage across this domain continuously and although they have not risen to the level of a declared war, the outcome of some of these battles could have just as significant of an impact on the long term future of the states involved in these ongoing conflicts [2]. Critical computing resources are captured, industrial and military secrets are stolen, strategic plans and diplomatic negotiations are compromised and key government, private, military and infrastructure systems are infiltrated, all to gain a competitive advantage for the states initiating these attacks.

The methods and processes employed to attack and defend information resources in cyberspace constitute maneuver as they are undertaken to give one actor a competitive advantage over another. As various nation-states throughout the world have begun building cyber warfare programs and have actively begun conducting operations in cyberspace, it is important to understand what constitutes the principle of maneuver in cyberspace as it relates to the traditional concept of maneuver in warfare and how the unique characteristics of the cyberspace warfighting domain alter this concept.

2. MANEUVER AS A PRINCIPLE OF WAR

The principle of maneuver has evolved as a tenant of war over the course of several thousand years. Beginning in the earliest recorded battles, the concept of maneuver involved the movement of troops to positions of advantage to attempt to fix and destroy enemy forces. Early forms of engagement included maneuvers such as the single envelopment, the double envelopment and the penetration and were mostly tactical in nature. As technology evolved, commanders were able to leverage new forms of transportation to increase the speed and tempo of maneuver in battle. Additionally, advances in weapons technology introduced the concept of fires and altered the principle of maneuver. At this point the use of maneuver came to resemble more modern definitions of employing forces through movement in combination with fires but was still largely tactical in nature.

The 1700s and 1800s saw the rise of operational maneuver as Napoleon's Grand Army swept through Europe in 1805 [3]. While Napoleon recognized and utilized operational maneuver, it was not until the battles of the American Civil War that it truly became institutionalized as a formal part of doctrine [4]. During World War Two, the German's use of Blitzkrieg ushered in another evolutionary step in maneuver shifting from attrition to maneuver warfare. Prior to

World War II, maneuver focused almost exclusively on destroying or defeating the enemy and sought to engage the enemy in decisive battles. By attacking though Belgium and avoiding the strength of the French army, German armored formations were able to drive deep into the enemy rear to achieve strategic success. "The effects of the lightening deep penetrations created a state of paralysis on the French military command forcing the capitulation of France itself" [5]. The development of Blitzkrieg by the Germans and similar developments in other militaries led to the concept of maneuver warfare which focused on incapacitating the enemy through shock and disruption rather than through attrition warfare.

During the 1970s and 80s, Colonel John Boyd developed theories which described maneuver in terms of competitive decision cycles. According to Boyd, "Victory in competitive decision cycles requires one side to understand what is happening and act faster than the other" [6]. Boyd's theories again revolutionized the principle of maneuver as they focused on creating the ability to make appropriate decisions faster than an opponent rather than on kinetic movement and fires. Maneuver in Boyd's terms could be described as "to operate inside an adversary's observation-orientation-decision-action (OODA) loops or get inside his mind-time-space to penetrate an adversary's moral-mental-physical being in order to isolate him from his allies, pull him apart and destroy his will to resist" [7]. Boyd was a key designer of the strategy the United States used to decisively defeat Iraq in the first Gulf War, the asymmetric success of which shocked many other states and led to what was called a Revolution in Military Affairs.

Modern definitions of maneuver owe a great deal to Boyd and many other military theorists and are an amalgamation of the experience of generations of military strategists. This discussion has very briefly described how the principle of maneuver has evolved and has necessarily skipped many important theorist and contributing theories in favor of brevity. Entire books could be written on how these theories have evolved over time but that is outside the scope of this paper. For purposes of this discussion, it is important to understand that "the essence of maneuver is taking action to generate and exploit some form of advantage over the enemy" [8]. Distilled down to its most basic form, maneuver can be simply defined as movement towards an objective. With this understanding in mind, it is appropriate to attempt to understand how the principle of maneuver applies to the domain of cyberspace and how the unique characteristics of this domain alter this concept.

3. CYBER MANEUVER

Cyber Maneuver is the application of force to capture, disrupt, deny, degrade, destroy or manipulate computing and information resources in order to achieve a position of advantage in respect to competitors. Maneuver in the traditional warfighting domains primarily involves the movement of military forces and application of fires, however, in cyberspace, there is obviously no movement of forces in the kinetic sense since it is a virtualized environment. Instead, maneuver in cyberspace involves the application of force to specific points of attack or defense. This force is the special purpose code written to accomplish the attacker's or defender's objectives and is implemented at the time and virtual location of their choosing. In a very real sense, forces do not move in cyberspace, the point(s) of attack are moved [9]. This makes observation and detection very difficult, especially in relation to the source of attacks.

Cyber maneuver is used to influence human and machine behavior. In a certain sense that is a redundant statement since the purpose of influencing machine behavior is ultimately to influence human behavior. Cyber maneuver leverages positioning in the cyberspace domain to disrupt, deny degrade, destroy or manipulate computing and information resources. It is used to apply force, deny operation of or gain access to key information stores or strategically valuable systems.

Another key factor in considering maneuver in cyberspace is that thus far, there has not been any open, state-on-state, cyber wars. There is, however, a constant state of conflict between states, surrogates or proxies, non-state actors and private entities and a great deal of evidence exists pointing to state involvement in much of this ongoing conflict. It is therefore advantageous to consider not just enemy states, but other adversaries and competitors when describing maneuver in cyber operations. International laws are still relatively immature in regard to cyber warfare, and so long as that remains the case, it is very likely that states will leverage this ambiguity to take actions in cyberspace that would be unacceptable in the physical world.

In defining cyber maneuver, it is important to understand the characteristics that make maneuver in cyberspace unique and to try to identify the major forms of both offensive and defensive maneuver that have thus far emerged in this domain. It should be noted that this effort is not meant to be exhaustive or all-inclusive. This is merely a starting point to try to quantify the trends that are emerging in this relatively new warfighting domain and to provide a basis for others to continue to refine doctrine in relation to cyberspace operations.

A. Characteristics of Cyber Maneuver

Cyberspace is a unique environment comprised of physical, informational and cognitive elements that blend together to create the virtual domain across which cyber operations occur. The principle of maneuver, when applied to operations in cyberspace, has distinct characteristics when compared to maneuver associated with the other warfighting domains of air, land, sea and space.

1. Speed

One of the most obvious characteristics of maneuver is cyberspace is the speed at which it can occur. Actions in cyberspace can be virtually instantaneous, happening at machine speeds. The speed at which actions can take place in cyberspace makes it incredibly difficult for one actor to react and adjust to a successful attack or to the modification of a defensive formation. By the time a successful attack is detected and mitigation undertaken, it is likely that either data has already been compromised or worse, hostile actions have already been completed to the detriment of the defending unit. If a modification is made to an element's defense in the midst of an attack, it is unlikely the attacker will be able to modify the attack quickly enough to continue successfully without being detected. In cyber operations, speed favors the side which has gained the initiative and successful maneuver allows an attacker or defender to get inside their adversaries' decision cycles and move more rapidly than they can react. Speed is a double edged sword in cyberspace. Actions happen at machine speeds, but reactions tend to happen at human speeds since reactions usually require some form of analysis and the involvement of a decision maker.

2. Operational Reach

Maneuver in cyberspace has almost unlimited operational reach. “Operational reach is the distance over which military power can be concentrated and employed decisively” [10]. In kinetic operations, operational reach is limited by terrain and distance, but since distance is virtually meaningless in cyberspace, reach in cyber operations tends to be limited by the scale of maneuver and the ability of an element to shield its actions from enemy observation, detection and reaction.

3. Access and Control

Maneuver in cyberspace requires access to friendly, neutral and enemy systems and one of the main goals of maneuver in cyberspace is to gain access to these systems in order to facilitate follow-on operations such as exploitation of data, disruption of systems or to gain leverage. Gaining control of systems is synonymous with building forward bases in a kinetic operation. It allows an attacker to move the point of attack forward to systems that are not attributable to the initiating state and potentially escalates an attacker’s privilege level relative to the ultimate target system or network.

4. Dynamic Evolution

The technology upon which cyberspace is based is constantly evolving. Recent years have seen rise to heavy use of web based applications, cloud computing, smart phones, and converging technologies. This ongoing evolution leads to constant changes in tactics, techniques and procedures used by both attackers and defenders in cyberspace. Methods that work today may not work tomorrow due to new and unforeseen technological advances. Unlike kinetic conflicts the battlefield terrain can shift presenting very little room for planning. Surveillance of the targets and defences can offer an advantage.

5. Stealth & Limited Attribution

Stealth and limited attribution have become the hallmarks of most attacks in cyberspace. Cyberspace is dominated by non-state, bad actors and sophisticated state actors that use the advantage of anonymity to mask their actions, making them unattributed [11]. Even large scale, overt attacks such as distributed denial of service (DDoS) attacks are most often difficult to attribute to a specific actor or state.

Every action that takes place in cyberspace is observable at some level. That being said, most actions are not observed in a meaningful way. This may be due limited sensor coverage, limited analysis capability or a number of other factors and it is these factors that assist attackers in hiding their attacks. Additionally, the ability to leapfrog from compromised system to compromised system makes attribution very difficult, especially when the systems in question are geographically dispersed in different international jurisdictions.

6. Rapid Concentration

In cyber space, attacks can rapidly build from a single source system to thousands or even tens of thousands of systems with little or no warning to the target system. In kinetic operations, it is very difficult for an attacker to generate this type of mass with little or no warning, especially in the modern era of satellite imagery, radar, etc. In cyberspace, attackers can make use of botnets

and crowd-sourcing to rapidly generate distributed mass effects that are especially effective in attacks like distributed denial of service attacks. This type of massing can also be used to hide more subtle attacks, distracting defenders who are attempting to restore services from these massed attacks while attackers conduct more covert penetration attacks.

7. Non-serial and Distributed

Maneuver in cyberspace allows attackers and defenders to simultaneously conduct actions across multiple systems at multiple levels of warfare. For defenders, this can mean hardening multiple systems simultaneously when new threats are discovered, killing multiple access points during attacks, collecting and correlating data from multiple sensors in parallel or other defensive actions. For attackers, this can mean simultaneously attacking numerous targets at multiple locations in parallel rather than engaging in serial attacks. "Serial attack is the old fashioned ebb and flow of battle. It is a linear concept where two adversaries engage in a series of attacks and counter attacks. In parallel attack, the point of attack is against multiple targets and the effects are non-linear" [12]. These non-linear effects can create serious dilemmas for defending units who often have limited resources to defend large numbers of systems. This is especially true when attackers focus their attacks at multiple levels generating tactical, operational and strategic effects simultaneously.

B. Basic Forms of Offensive Cyber Maneuver

Cyber Maneuver most differs from its kinetic counterparts in offensive operations. While the goal of maneuver, to secure positional advantages in respect to an enemy or competitor state, remains relatively consistent with kinetic maneuver, the means to do so is vastly different given that maneuver is conducted at machine speeds inside a virtual construct.

1. Exploitive Maneuver

Exploitive Maneuver is the process of capturing information resources in order to gain a strategic, operational or tactical competitive advantage. It is modern day espionage at its finest, but it is the use of this information in follow-on operations that makes it a valid and dangerous form of cyber maneuver. In this new warfighting domain, information is analogous to terrain and the capture of key information resources can lead to decisive results across the political, economic, financial or military spectrums. Unlike terrain on a kinetic battlefield, once captured, information resources cannot be retaken to regain an advantage. On the kinetic battlefield, a key piece of terrain captured by the enemy can potentially be counter-attacked and the advantage of holding that terrain regained for future operations. This is not true in the information environment when dealing with sensitive data or information stores. Once critical information resources are exposed, the originating state often loses a significant competitive advantage and the gaining state utilizes these resources for its own purposes.

Over the course of the last decade, various nation-states have recognized the competitive advantage they can gain by harvesting the intellectual property and state secrets of competitor nations. Chief among these has been China which has been conducting large scale cyber operations to capture information resources. "China has made industrial espionage an integral part of its economic policy, stealing company secrets to help it leapfrog over U.S. and other foreign competitors to further its goal of becoming the world's largest economy"

[13]. Additionally, there is some anecdotal evidence that suggests that China has used captured information resources to give it distinct advantages when engaging in diplomatic or corporate negotiations. A recent investigation in Canada linked Chinese hackers to intrusions at several law firms and government offices in an apparent effort to gain a strategic advantage in ongoing deal negotiations. "The investigation linked the intrusions to a Chinese effort to scuttle the takeover of Potash Corp. of Saskatchewan Inc. by BHP Billiton Ltd. as part of the global competition for natural resources" [14].

While espionage is certainly not new, cyberspace has enabled the capture and exploitation of information on an unprecedented scale. Given that information is analogous to terrain in cyberspace, it stands to reason that the processes involved in attacking and defending it must represent a key form of maneuver in cyber operations.

2. Positional Maneuver

Positional Maneuver is the process of capturing or compromising key physical or logical nodes in the information environment which can then be leveraged during follow-on operations. These nodes could be viewed as centers of gravity in the information environment and gaining logical control of these nodes will give the attacker key advantages and leverage during the escalation of conflict, especially in the case of war or other combat operations. "Leverage is used to impose a force's will on the enemy, increase the enemy's dilemma, and maintain the initiative" [15]. The logical nodes in question could be Supervisory Control and Data Acquisition (SCADA) systems, enemy command and control systems, systems designed to provide a common operational picture during combat operations or any other key system whose compromise at a key moment in battle could give the initiating force a decisive advantage.

A prime example of this kind of positional maneuvering could be intuited from the 2007 Israeli attack on a suspected nuclear reactor at Dayr az-Zawr, Syria. Israeli strike aircraft managed to fly into Syria without alerting Syrian air defense systems to carry out this raid. This was apparently accomplished through a combination of both electronic and cyber-attacks which caused all of Syria's air defense radar systems to go offline for the duration of the raid [16]. Before the kinetic operation could be undertaken, the Israelis had to know that they could disrupt the systems in question. This implies that the Israelis had already gained the necessary level of access into these systems and had pre-positioned themselves to carry out this attack. They had to be confident they could disrupt these critical systems at the time of their choosing to ensure the success of the raid. The use of positional maneuver prior to the initiation of actual kinetic combat operations set them up for success and illustrates the potential decisive nature of this form of cyber maneuver, especially at the tactical and operational levels of war.

3. Infleuncing Maneuver

Influencing Maneuver is the process of using cyber operations to get inside an enemy's decision cycle or even to force that decision cycle though direct or indirect actions. This is a broad form of maneuver intended to gain and maintain information superiority and dominance and to maintain freedom of maneuver in cyberspace. Influencing maneuver is often used in conjunction with other forms of offensive maneuver. Influencing maneuver can be used in direct or indirect operations. A direct example of influencing maneuver could include actions

such as compromising command and control systems and manipulating data subtly in order to degrade the confidence a commander has in his systems to slow down his decision cycles. Indirect actions might include feeding compromised and manipulated data to the media to force a desirable reaction from an enemy. Influencing maneuver falls heavily in the spectrum of traditional information operations but makes use of cyber maneuver to accomplish its objectives.

C. Basic Forms of Defensive Cyber Maneuver

To date, defensive maneuver in cyberspace generally resembles its kinetic counterparts. Perimeter defences, intrusion detection, and defense-in-depth is almost identical in concept whether executed in a kinetic defense or in the virtual world of cyberspace and the Deceptive Defense is somewhat akin to an ambush, luring in an attacker although for somewhat different purposes. The Moving Target Defense is unique to the cyberspace and relies on technical mechanisms that do not have a true analogy in the physical world.

Cyber defense is often seen as being much more difficult than offensive operations due to what is perceived as an asymmetric advantage on the side of the attacker. While that is largely true, the proper use of defensive maneuver can offset that advantage and allow defenders to regain the initiative. "Cyber defense seeks to anticipate and avoid threats, detect and defeat threats, survive and recover from attacks. In an analogy to the OODA loop, cyber defense seeks to operate inside the OODA loop of the threat" [17].

1. Perimeter Defense & Defense in Depth

Line Defense is the Maginot Line of cyberspace and like this historic example; it is highly susceptible to maneuver. The line defense is used by many organizations who spend resources protecting the perimeter of their network with firewalls, intrusion detection systems and other defensive measures but leave the interior of their networks relatively undefended. Defense in depth is mitigation strategy that attempts to mitigate the vulnerabilities of the line defense by hardening the interior of the network and individual systems as well. While defense in depth is a more effective strategy than a line defense, both these defensive formations suffer from the fact that they are fixed targets with relatively static defenses which an enemy can spend time and resources probing for vulnerabilities with little or no threat of retaliation.

2. Moving Target Defense

The Moving Target Defense, unlike the line defense discussed above, does not attempt to create impenetrable defensive rings to prevent attacks and protect resources. Instead, this form of defensive maneuver uses technical mechanisms to constantly shift certain aspects of targeted systems to make it much more difficult for an attacker to be able to identify, target and successfully attack a target. A Moving Target Defense attempts to "create, evaluate and deploy mechanisms and strategies that are diverse, continually shift, and change over time to increase complexity and costs for attacker, limit the exposure of vulnerabilities and opportunities for attack, and increase system resiliency" [18]. Typically, Moving Target Defenses use one of three methods, Address Space Randomization, Instruction Set Randomization and Data Randomization, to attempt to thwart attacks although other forms of system diversification are currently being researched.

During the 2008 cyber-attacks against Georgia, the Georgian government demonstrated a rudimentary form of the Moving Target Defense by relocating its primary sites on servers in several other allied countries. "The Georgian government took an unorthodox step and sought cyberrefuge in the U.S., Poland and Estonia. Within the U.S., Georgia located its cybercapabilities on servers at Tulip Systems (TSHost) in Atlanta, Ga., and at Google in California. When Estonia experienced a cyberattack in 2007, it essentially defended in place; Georgia, on the other hand, maneuvered" [19]. By employing defensive maneuver, Georgia was able to maintain key government services in the face of a massive denial of service attack which was largely successful against its original Defense-in-Depth strategy.

3. Deceptive Defense

Deceptive maneuver is the cyberspace analogy to an ambush. Deceptive maneuver uses processes to lure an attacker in to committing actions which will reveal their methodology or assist the defender in attribution. An excellent example of this is the use of honeypots, purposely vulnerable systems designed to appeal to an attacker as an attractive target. The use of these types of systems can allow a defender to regain the initiative by stalling an attack, giving the defender time to gather information on the attack methodology and then adjusting other defensive systems to account for the attacker's tactics, techniques and procedures.

4. Counter Attack

The counter attack is another form of defensive maneuver and has a direct kinetic counterpart. While the concept of a counter attack is relatively straight forward, the execution of a counter attack in cyberspace is complicated by the difficulty of attribution and the fact that many attacks originate from compromised, third party systems. Taking these issues into account, counter attacks may prove necessary to restore critical operations even at the cost of disabling or damaging a compromised third party system. In situations where attribution has been established, the use of a counter attack can allow a defender to stall an attack and regain the initiative. Consider a situation in which the command and control server for a botnet has been identified. Conducting a counter attack against such a system could disrupt a distributed attack and allow the defender to restore operations.

4. SOVEREIGNTY ISSUES AND CYBER MANEUVER

Sovereignty can be defined as a state exercising authority and control over a given area or geographic region. In relation to sovereignty, cyberspace is informally considered a global commons, similar to the sea and air domains, in that it is considered to be outside the geographic jurisdiction of any particular state and is an internationally shared resource utilized for trade, communications and other uses. Cyberspace is also described as a borderless domain, but that is not an entirely accurate statement and there are a number of different means that states can and do use to justify sovereign control of portions of this domain.

A. Efforts to Define Borders in Cyberspace

A number of states such as China have begun filtering content at the logical borders of their portion of cyberspace and in doing so have created de facto borders by exercising control

and authority over these virtual regions. Additionally, a number of states including the United States are currently exploring policies on how to define national borders in cyberspace [20]. This makes sense in both political and military contexts since it is currently difficult to cry foul for virtual incursions when there is no formal policy defining what the United States considers to be its sovereign territory in this domain. However, individual states defining sovereignty in cyberspace have limited utility without international agreements acknowledging the right to sovereignty in this domain. Both the United States and Russia have publically declared that they reserve the right to respond to cyber-attacks using all means at their disposal to include traditional kinetic options. This implies that the current state of this issue is based more on right-by-might than any form of international consensus.

One difficulty in defining borders in cyberspace is that the physical geography of cyberspace does not even remotely match the logical geography. Every router, switch and device upon which the domain of cyberspace exists is physically located within a state. One could use this as an argument to use state borders as a map of cyber borders. In this model, all systems residing inside the United States and its territories would be considered to be with the sovereign control of the United States and attacks on these systems would represent a violation of that sovereignty and a hostile act. While this may seem like a simple and straight forward way to deal with this issue, it would leave many US systems unprotected when you consider the logical borders of US systems in cyberspace. "The United States Military operates a global, logical domain (Dot MIL) that spans over 88 countries in over 3,500 locations. This logical domain interconnects with more than 20,000 leased circuits and supports over 2.8 million users" [21]. Clearly the United States would consider an attack by a competitor state against its military systems, even those residing outside the United States, to be a hostile act. Therefore, simply relying on physical boundaries does not fully address the issue. However, the complexity involved in trying to establish logical borders is insurmountable. "There is no clear-cut way to establish a permanent or even semi-permanent cyberspace boundary using the logical boundary approach. The demarcation point would be in a constant state of fluctuation" [22]. Even with the current ambiguity over sovereignty in cyberspace, there are forms of cyber maneuver that could still be considered hostile acts and violations of sovereignty.

B. Violating Sovereignty in Cyberspace

Viewed in its current state, cyberspace resembles a vast frontier with millions of small enclaves, many of which are surrounded by defensive perimeters. While the Internet is sometimes described as borderless, this is more of a legal distinction involving "jurisdictional uncertainty and transcendence of international borders" [23]. In reality, the electronic perimeters of various enclaves do provide a version of borders that, in reference to maneuver, could have significant importance. It would be easy for a state to claim a violation of its sovereignty based on a cyber-attack on these enclaves, especially when these enclaves represent government or military organizations. The state in question has a vested interest in protecting these enclaves, and is exercising control and authority over them.

Consider a state which exercises positional maneuver to put a Remote Access Tool (RAT) into another state's SCADA systems, especially systems associated with critical infrastructure. While this action has not technically damaged these systems, the presence of this tool suggests

a future intent to make illicit use of it in what might be a very damaging attack. This could be construed as a precursor to a first strike. Additionally consider if a state like Iran used exploitive maneuver to capture information on nuclear weapons technology from Israel or the United States. Such maneuver could easily trigger a kinetic response given the public policy these states have against allowing Iran to gain nuclear weapons. While the above examples are both fictitious, both illustrate how actions in this domain could be seen as violations of a state's sovereignty.

Another serious consideration in regards to sovereignty and cyber maneuver is the concept of neutral states. In kinetic operations, a state must generally get permission from another state if its maneuver will cross that state's physical borders. How does this translate to the cyberspace domain when virtually any action between states will involve crossing national, international, state and non-state boundaries on both the physical and logical levels? Additionally, maneuver and attacks often involve the use of third party, neutral systems to mask attribution and provide the initiating state plausible deniability for the actions it initiates. Translating this to an example in the physical world, imagine what the United States' response would be if Canada somehow managed to fire missiles at Mexico from Texas. Yet events like this happen constantly in the cyberspace domain and rely on stealth and limited attribution to avoid political recriminations.

As more states begin to explore the idea of sovereignty in cyberspace, its relevance to cyber maneuver will continue to grow in importance. However, until some consensus is reached in the international community as to whether there exists a right to sovereignty in cyberspace and on what basis borders will be defined; this will remain an area of ambiguity that can be exploited in cyber operations.

5. ANALYSIS AND CONCLUSIONS

The principle of maneuver remains an important warfighting principle in cyberspace, but there are significant differences that must be taken into account when defining this concept. Information is the currency of warfare in cyberspace. Maneuver is used in cyberspace to position and apply force to attack or defend information resources much as kinetic maneuver makes use of key terrain in the physical world. Unlike terrain however, the capture of information resources can have a much more lasting impact at all levels of engagement since once exposed, the value of information depends on its usefulness to both the attacker and defender. This value can represent a short term gain such as exposure of tactical plans, or could have an impact that spans years such as the exposure of highly classified technologies.

Like its kinetic counterpart, cyber maneuver is used to give an actor a position of advantage over its enemies. Unlike kinetic maneuver, it is also highly applicable to adversaries and competitor states, even if those states are political allies. Cyber operations have not been limited to enemy states battling each other. Allied states with competing economic and political agendas are undoubted using these tools to secure competitive advantages. Proper use of cyber maneuver allows a force to maintain freedom of action in the cyberspace domain and can lead to competitive advantages in economic, political and military strategies.

Initiative is vitally important to cyber maneuver since actions are far quicker than reactions in this domain. Losing the initiative in cyberspace can leave a force paralyzed as it tries to apply human analysis and decisions to actions that are happening at machine speeds. Unlimited operational reach combined with non-linear effect compound this issue and add to the complexity faced by decision makers when reacting to enemy maneuvers.

Sovereignty issues will play an important role in cyber maneuver as various states and the international community try to come to some consensus on whether the concept of borders are applicable to cyberspace and if so, how to define them. Current difficulties in determining attribution for attacks combined with legal ambiguity make it advantageous for attackers to operate outside their parent state's sovereign systems. Attackers have a vested interest in dispersing attack sources; however, this could potentially present some significant issues since it involves launching attacks from systems belonging to enemy, neutral, or even allied third parties. So long as the current status quo remains, this type of attack pattern will probably remain prevalent and cyber maneuver will take this into account. Should attribution become easier due to technology changes, or should the international community come to terms with sovereignty issues in cyberspace, this could lead to significant changes in how maneuver is conducted in cyberspace, especially in regard to use of third party systems as jump off points for attacks.

One of the most dominant characteristics of maneuver in cyberspace is the fact that blatantly hostile acts are often accomplished with little or no recrimination against the initiator due to anonymity and the difficult of attribution. In many cases, similar acts in the physical world would be easily considered acts of war. Consider the Stuxnet virus which is thought to have disabled or damaged approximately 1000 centrifuges at the Natanz Nuclear Facility in Iran [24]. Outside of accusations in the media, Iran has done little in the way of retribution for this attack. Had this attack been carried out kinetically, it is very likely that it would have resulted in retaliation against the initiating state although what form that retaliation would have taken remains open to debate.

As states around the world are building and developing cyber warfare programs, understanding how the principles of war apply to this new warfighting domain becomes increasingly important since it is these principles that strategists and theorists use to develop strategy and doctrine. Maneuver has a critical role in this doctrine since maneuver is an integral tool that supports and enables other warfighting functions and principles. Maneuver is used to build mass, bypass strength, exploit vulnerability, gain and maintain the initiative and exploit success to achieve a state's tactical, operational and strategic objectives. While maneuver in cyberspace is uniquely different than its kinetic counterparts, its objective remains the same, to gain a position of advantage over a competitor and to leverage that position for decisive success. It is therefore important to continue to study and define the evolving principle of maneuver in cyberspace to ensure the success of operations in this new warfighting domain.

ACKNOWLEDGEMENTS

The author would like to gratefully acknowledge the efforts of Dr. Angelos Stavrou of George

Mason University and Major André Abadie of the United States Army who assisted in the editorial review of this paper.

REFERENCES:

[1] Joint Publication 3-0: Joint Operations, JP 3-0. Joint Chiefs of Staff, United States Department of Defense, Washington D.C., 2011, p. III-27.

[2] J. Markoff, "Before the Bunfire, Cyberattacks," The New York Times, Aug. 12, 2008; http://www.nytimes.com/2008/08/13/technology/13cyber.html.

[3] J.N. Wasson, "Innovator or Imitator: Napoleon's Operational Concepts and the Legacies of Bourcet and Guibert," SAMS Monograph, School of Advanced Military Studies, Command and General Staff College, Fort Leavenworth, KS, 1998.

[4] M.J. Lyons, "Napoleon, 'Stonewall' Jackson, and Operational Art," Desaxx – Military History – Military Art – National Security, Sep. 7, 2010, http://desaxx.blogspot.com/2010/09/napoleon-stonewall-jackson-and.html.

[5] F. Zachar, "Strategic Maneuver: Defined for the Future Army," SAMS Monograph, School of Advanced Military Studies, Command and General Staff College, Fort Leavenworth, KS, 2000; http://cgsc.cdmhost.com/cdm/singleitem/collection/p4013coll3/id/576/rec/8.

[6] L. Wells II., "Maneuver in the Global Commons – The Cyber Dimension," SIGNAL Magazine, Dec. 2010; http://www.afcea.org/signal/articles/templates/Signal_Article_Template.asp?articleid=2472&zoneid=306.

[7] J.R. Boyd, "Patterns of Conflict," Project on Government Oversight, Defense and National Interest – John Boyd Compendium, 1986; http://dnipogo.org/john-r-boyd/.

[8] Marine Corps Doctrinal Publication 1-0: Marine Corps Operations, MCDP 1-0, Department of the Navy Headquarters, United States Marine Corps, Washington D.C., p. 6-3.

[9] R.C. Parks & D.P. Duggan, "Principles of Cyber Warfare," IEEE Security & Privacy, vol. 9, no. 5, p. 31, Sep./Oct. 2011.

[10] Joint Publication 3-0: Joint Operations, JP 3-0. Joint Chiefs of Staff, United States Department of Defense, Washington D.C., 2011, p. III-28.

[11] J.R. Schilling, "Defining Our National Cyberspace Boundaries," Masters Thesis, National War College, Washington D.C., 2010.

[12] P.K. Singh, "Maneuver in Cyberspace," MMS Thesis, Command and Staff College, Marine Corps University, Quantico, VA.

[13] M. Riley & J Walcott, "China-Based Hacking of 760 Companies Shows Cyber Cold War," Bloomberg, Dec. 14, 2011; http://mobile.bloomberg.com/news/2011-12-13/china-based-hacking-of-760-companies-reflects-undeclared-global-cyber-war?category=%2Fnews%2Findustries%2F.

[14] M. Riley & S. Pearson, "China-Based Hackers Target Law Firms to Get Secret Deal Data," Bloomberg, Jan. 31, 2012; http://www.bloomberg.com/news/2012-01-31/china-based-hackers-target-law-firms.html.

[15] Overview of Operational Art, Joint Electronic Library, Defense Technical Information Center, Joint Doctrine Reference Materials, [Online] Available: www.dtic.mil/doctrine/jrm/opart.doc.

[16] D.A. Fulghum & R. Wall, "U.S. Electronic Surveillance Monitored Israeli Attack on Syria," Aviation Week, Nov. 14, 2007; http://www.aviationweek.com/aw/generic/story_generic.jsp?channel=defense&id=news/ISRA112107.xml&headline=U.S.%20Electronic%20Surveillance%20Monitored%20Israeli%20Attack%20On%20Syria.

[17] K.T. Jabbour, "50 Cyber Questions Every Airman Can Answer," Wright Patterson Air Force Base Public Affairs, Wright Patterson Air Force Base, Ohio, 2008; http://www.au.af.mil/au/awc/awcgate/afrl/50_cyber_questions.pdf.

[18] S. Jajodia, A.K. Ghosh, V. Swarup, C. Wang & X.S. Wang, Eds. New York, Springer Science + Business Media, 2011.

[19] S.W. Korns, "Botnets outmaneuvered," Armed Forces Journal, Jan. 2009; http://www.armedforcesjournal.com/2009/01/3801084/.

[20] J.R. Schilling, "Defining Our National Cyberspace Boundaries," Masters Thesis, National War College, Washington D.C., 2010.

[21] J.R. Schilling, "Defining Our National Cyberspace Boundaries," Masters Thesis, National War College, Washington D.C., 2010.

[22] J.R. Schilling, "Defining Our National Cyberspace Boundaries," Masters Thesis, National War College, Washington D.C., 2010.

[23] K.T. Jabbour, "50 Cyber Questions Every Airman Can Answer," Wright Patterson Air Force Base Public Affairs, Wright Patterson Air Force Base, Ohio, 2008.

[24] D. Albright, P. Brannan & C. Walrond, "Did Stuxnet Take Out 1,000 Centrifuges at the Natanz Enrichment Plant?" Institute for Science and International Security, Washington, D.C., Dec. 22, 2010; http://isis-online.org/uploads/isis-reports/documents/stuxnet_FEP_22Dec2010.pdf.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Countering the Offensive Advantage in Cyberspace: An Integrated Defensive Strategy

David T. Fahrenkrug
Office of Net Assessment
Office of the Secretary of Defense
Arlington, VA, USA
david.fahrenkrug@gmail.com

Abstract: Current accepted wisdom in cyberspace is that the attacker has the decisive advantage. The number of detected intrusions across public and private networks is increasing at an alarming rate, while the costs to defend against these intrusions are rising exponentially. Today's best cyber security costs nearly ten times as much as the malware it is designed to protect against. This strategy is unsustainable. Drawing from defensive strategies used in other domains, this paper will offer an integrated defensive strategy for cyberspace that could even yield a decisive advantage over the offense.
An integrated defense begins by first trying to avoid the attack by actively dispersing the networks and information using IP and frequency hopping, data fractioning, cloud dispersal, and steganography. Second, an integrated defense includes hardening the infrastructure and date using encryption and shielding of electronic components. Finally, an integrated defense is able to detect and respond to intrusions and attacks. This requires an accurate and continuously updated awareness of the network's configuration and activity as well as the ability to recover and respond to the attack.

Keywords: *defense, maneuver, dispersal, encryption, hardening, detection*

1. INTRODUCTION

On the morning of Nov 17, 1917, the British commenced an attack against the Germans in a little town of Cambrai. This battle marked the first time tanks, artillery, infantry, and aircraft were combined in a coordinated, synchronized campaign to outmaneuver the heavily fortified defenses of World War I trench warfare. Twenty-two years later, the Germans used those same technologies and capabilities to sweep across Europe with a revolutionary concept of warfare they referred to as *Blitzkrieg*. By organizing these very different combat arms into a combined

The views presented in this paper are of the author and not necessarily representative of the view or policies of the Department of Defense or the Office of Net Assessment.

form of maneuver warfare, the Germans were able to defeat the most sophisticated—and expensive—defensive system in the world, the French *Maginot Line*. The failure of the *Maginot Line* to withstand the German attack was primarily the result of a static defensive strategy that did not anticipate the speed of maneuver *Blitzkrieg* would be able to achieve on the battlefield.

Today's current cyber defenses suffer from a similar lack of flexibility and maneuverability. Like the *Maginot Line*, today's cyber defenses are not failing due to a lack of new technologies. In fact, sufficient capability and technology exist today to counter and possibly reverse the advantage of the attackers. Instead, today's cyber defenses are failing because they lack the organizing concepts that can integrate current capabilities into a flexible and adaptive strategy. From a military point of view, the ability to organize and integrate capabilities to achieve specific objectives is known as the operational art of war. Commanders and operational planners bring together various capabilities and tactics and integrate them into lines of operation designed to achieve specific operational objectives that ultimately contribute to the overall campaign strategy. Drawing from defensive strategies used in other domains, this paper will offer an integrated defensive strategy for cyberspace.

The first section of this paper provides a description of cyberspace that will become the basis for crafting a defensive strategy. The next section will then review defensive concepts from other domains and introduces four principles of an integrated defensive strategy. The remaining sections will then apply these four principles to cyberspace to illustrate how an integrated cyber defense could be implemented. The paper concludes with a brief discussion on the critical next steps that should be pursued.

2. CYBERSPACE

Before introducing new ways to improve the defense, we must first understand what we are defending, why we are defending it, and where we are defending it. Even though cyber is most often used as a metaphor for the internet, computers, or hacking in general, a more useful understanding of cyberspace is reflected in the United States Department of Defense definition [1]. "Cyberspace is a global domain within the information environment consisting of the interdependent network of information technology infrastructures, including the Internet, telecommunications networks, computer systems, and embedded processors and controllers". From this definition, we see that cyberspace is a variety of networked systems that were created by connecting electronic components using signals (electromagnetic energy) and software. More importantly, cyberspace was created so that we could more easily and rapidly create, store, modify, and transfer data and information. This description of cyberspace allows us to distinguish the place—cyberspace—from the activities that occur within that place. The principle roadblock to gaining greater understanding of warfare and competition in cyberspace has been simply conflating the networks with their functions. What we do with networks is fundamentally different than the network itself.

So, we must distinguish between cyberspace and how cyberspace is used. The pervasiveness of networks and the number of systems and functions that now rely on the rapid transfer of data is a testament to how important this new "terrain" has become. While networks vary significantly

from one another by the type of hardware, software, or signals that are used to create the connections, they all exist for essentially the same reason; to improve and increase our ability to transfer data and information. Thus, we are interested in not only protecting our ability to access and use cyberspace, but more importantly we want to protect the functions and data that are resident in cyberspace. These are not the same thing and protecting them may require very different approaches.

One way to increase our understanding of the difference between place and function is to draw from theoretical treatises concerning other domains. For example, Julian Corbett [2] offers an elegant theory on naval warfare to include a perspective on the sea domain. Due to the lack of differentiation between information and cyberspace within the literature on information warfare, a key principle of warfare has been misunderstood—control of the operational domain. Julian Corbett describes this as the "object" of warfare. Regardless of the domain, the object of warfare in that domain is the attainment of some level of control over access and use of the domain. Corbett describes the principle most clearly in his discussion of maritime strategy. "The object of naval warfare must always be directly or indirectly either to secure the command of the sea or to prevent the enemy from securing it." Other theories also describe a requirement for controlling the domain first and then using it to achieve other objectives. Giulio Douhet [3] identified achieving "command of the air" as the first priority and the reason a nation needed an independent air force. In modern military doctrine, and in particular Air Force doctrine, this principle is often described as superiority.

A significant difference, however, from the other domains is the flexibility of the terrain in cyberspace and the lack of requirement to defend *specific* terrain. In other words, cyber defenses are not bound by territory. Rather than defending a piece of territory or area of airspace, cyber defenses are concerned with protecting content and function. If organized, planned, and exercised properly, any compromised component of a network could be isolated and even discarded while the functions and data continue to exist in the remaining elements or are rerouted to new infrastructures. This means cyber defense can become just as agile as the offense. This unique characteristic of cyberspace should figure prominently in any integrated defensive strategy.

3. INTEGRATED AIR DEFENSES

In the period between the Battle of Cambrai and the deployment of *Blitzkrieg* warfare, the necessary technology had already been discovered. Yet only the Germans had adopted this new form of maneuver warfare. The innovation did not come from new technology, but from employing new concepts of operation that integrated existing technologies to achieve greater speed and agility on the battlefield. The Germans developed and practiced a combined arms approach to create synergy between the tank, infantry, air, and artillery components that resulted in a maneuver advantage that was difficult to overcome in the early years of World War II. In the same way, sufficient technologies exist today to overcome the offensive advantage that is overwhelming current defenses. As Paul Williams [4], executive director of security services for White Badger Security, confidently claimed when asked at a recent conference about Stuxnet, "There's absolutely no way it would have happened with just a reasonable dose of off-the-shelf

commercial technology." This "reasonable dose", however, needs to be employed using an integrated operational concept in order to be effective against a maneuvering adversary.

Similarly, the tragedy of Pearl Harbor was that we were not ready to fight through an attack. With no warning of attack, aircraft parked closely together on open taxiways, and aircrew not prepared to respond immediately, the Japanese easily and swiftly destroyed most of the combat capability located at Pearl Harbor. Since that time, militaries have responded to the potential devastation that could be suffered from an air attack by developing integrated air defenses. Everywhere in the world, countries with sufficient resources have built integrated defensive systems based on a layered and responsive approach. With the United States military, despite the fact that no bases, or ground forces for that matter, have come under attack by aircraft in more than fifty years, the Air Force still trains as if they will.

An integrated defense begins with radar capable of detecting the threat to potentially provide early warning and direct a response against the attacker. The defense uses these warning and detection systems to cue aircraft flying defensive combat air patrols as well as surface to air missiles to counter the incoming attack. In preparation for the possibility that at least one attacker will get through, buildings and aircraft shelters are hardened, and personnel are trained on how to conduct rapid runway repairs. In addition, aircraft, support equipment, and even the runways, are dispersed to increase the number of targets and decrease the likelihood that any single attack could wipe out all capabilities. Finally, aircrews are trained to scramble and get their aircraft airborne as soon as possible. Applied to cyberspace, this means developing network sensors, offensive responses, and protection and recovery procedures for critical data and operating systems. More importantly, this means exercising and training for the eventuality of an attack.

4. INTEGRATED CYBER DEFENSE

During World War II, a key objective of the Allies was to secure the transfer of critical parts and supplies. In the face of a persistent German campaign, this meant at times actually escorting some of the ships with cruisers and submarines. The U.S. Navy did not try to secure all the sea lanes, all the time. In fact, there were certain aspects of the ocean that the Germans had free access to all the way up to the coast of the United States. Not all data is critical, and not all networks need to be secured. The key is ensuring that the mission can be accomplished. This concept of mission assurance is gaining traction throughout the military, but there is still a lack of operational concepts [5]. The following sections will describe each aspect of an integrated operational concept to improve cyber defenses.

A. Dispersal

When considering how to disperse forces and capabilities, we must once again first identify what we are dispersing and distinguish that from where we are dispersing it. Some networks are purely functional and do not directly affect information, while other networks exist only to store data and information. In the first case, we want to disperse the *functions* of the network, while in the latter, we want to disperse the *data or information*. The purpose in both instances is to make targeting that much more difficult for an adversary.

When dispersing the network, all aspects of the network environment must be considered for dispersal. Operating systems can be dispersed as virtual machines within the network or outside the network to mitigate a software attack. Communication lines, both wired and wireless, can be dispersed by increasing the number of fiber lines available or by using a greater range of frequencies of electromagnetic energy to transmit the data. Hardware components can also be distributed across multiple platforms to reduce the possibility that any one system becomes a single point of failure for the entire network.

For example, the recent STUXNET case highlighted the vulnerability of SCADA devices with only one algorithm for controlling a critical process. Keith Stouffer, Joe Falco, and Karen Scarfone [6] suggest a possible solution is to disperse functionality within the integrated control device. "Maintaining functionality during adverse conditions involves designing the ICS [integrated control system] so that each critical component has a redundant counterpart. Additionally, if a component fails, it should fail in a manner that does not generate unnecessary traffic on the ICS or other networks, or does not cause another problem elsewhere, such as a cascading event." The objective is to build resilient and survivable control systems through automated sensors, pre-established algorithms, and defined responses.

Similarly, storing complete sets of data and information in a single location simplifies that attacker's problem and in some cases even singles out the information as being more important. Cloud storage solutions offer the possibility of hiding data and information by placing it in a noisier environment. Ken Sorrels [7] argues that we need to inventory the functions and content of the network and then segment them off into different areas based on characteristics like confidentiality, integrity and availability. "This keeps an entire system from being at risk when a certain zone is breached."

Just as camouflage and decoys are effective ways to disguise the location of physical targets, so the expanding number of storage solutions presents an opportunity to disperse and hide information and functions resident in the network. The ability to disperse also provides an added benefit of increasing confidence levels in the veracity of the information. The more the information is fractioned and dispersed, the less likely an adversary will be able to corrupt or deny access to all of that information. Again, this is where it is important to understand and prioritize the information on the network or the functions the network is supporting.

For some information, the content is more critical than how quickly it can be accessed, while other information is only useful at a specific time and moment. For example, a flight of F-22s connected by a tactical data link share situational and targeting data that is time sensitive and often very perishable. What is most critical is that the data is received on time and in the format that is required to complete the kill chain.[1] The more perishable the data becomes, the more important timely reception of the data becomes. This places less emphasis on securing the signal, and more importance on ensuring sufficient pathways to deliver the data.

Rather than transmit data across a single, highly encrypted frequency (or narrow band of frequencies) that simplifies the adversary's detection and jamming problem, the data link and the data being transmitted should be dispersed across a range of frequencies within the

1 The military has codified the chain of events required to acquire and target an adversary. The kill chain is summarized by the phrase "find, fix, track, target, engage, assess" or the acronym F2T2EA.

electromagnetic spectrum. This type of spectral agility was explicitly identified in a recent military report [8] that identified the requirement for "jam-resistance, low-probability-of-detection/ intercept, and cyber resilience in the increasingly congested spectrum environment and increasingly contested electronic warfare environment." This accomplishes two things: first, the likelihood that an adversary can detect and then target each of the signals is decreased; and second, the veracity of the data is increased because an adversary must intercept and alter each instance of the data that has been transmitted. A simple voting scheme that compares each of the transmissions of the data can be used to verify the validity of the information that the other aircraft is receiving. In this case, nothing about the network or the data has to be "secured" because the information is perishable and of little use beyond that instance of time. Instead, dispersal of the signal and the data preserves the *ability* to transmit and receive data with increased confidence that the data has not been compromised.

In other cases, the same flight of F-22s may be sharing positional information of the formation available on the same tactical data links that could compromise the mission if intercepted by an adversary. The challenge then becomes one of securing the information *and* ensuring its availability to all members of the flight. This will require some level of encryption of the data, but not necessarily for the network itself. The point, once again, is we have to first identify and prioritize the data and functions that are dependent on the network and then choose the most effective way to distribute them using a combination of hardware, signals, and software.

B. Hardening

In addition to dispersal, the functions of the network and the information resident in the network need to be hardened. Dispersal increases the probability of avoiding the attack, while hardening increases the probability of surviving the attack. Again, existing technology is available that could be used to decrease the likelihood of an attacker accessing a network or affecting the contents of the network. While public key encryption is increasingly being used, the use of hash and private key encryption for information stored and transmitted on their networks needs to increase as well. Rather than trying to secure every network or computer system, businesses and organizations need first to prioritize their networks and information and then apply appropriate levels of encryption to ensure operating systems, data, and automated commands are not compromised.

Available encryption practices and an extensive number of software solutions can radically reduce the vulnerability of data and operating systems. For example, IBM has developed a secure processor chip that protects the operating system from physical or software attacks with no known compromises after fours years and millions of chips operating word wide [9]. Despite this success, not all encryption methodologies will be perfect all the time. In the most sensitive networks, containing the most sensitive data, encrypting across all aspects of the network adds layers of defense that greatly compounds the attacker's problems. Potentially, encryption methods could also incorporate steganography to further hide or disguise data or software even while it is being hardened through encryption.

For some networks, the hardware will need to be protected against electronic attack or persistent intrusion sets. In both cases, there are current and emerging technologies which can increase the resiliency of chips, processors, and control devices from malicious attacks. During the

Cold War, electronic components that could potentially be exposed to an electromagnetic pulse following a nuclear detonation (i.e., navigation and communication components on a B-52 bomber) were specifically designed to survive such a situation. In an effort to increase the speed of our chips and processors, these components have become even more vulnerable to some type of electromagnetic inference. While costs and weight clearly prohibit the hardening of every component within a network, there are ways to harden the most critical components.

When combined with dispersal, the chances that an attacker will be able to affect the data or the functions of the network are significantly reduced. In fact, the cost and time required to attack networks configured with these defenses will likely deter most potential attackers. Still, a determined adversary will get through eventually, such that an integrated defense must have the ability to detect the intruder and then respond.

C. Detection

At no time in the history of warfare has any commander had perfect awareness of the battlespace. Despite our best efforts to gain "information dominance" it will always elude us. Fog, friction, and uncertainty are fundamental characteristics of war that we may be able to mitigate in some circumstances, but never completely eliminate. Instead, our objective is to anticipate and prioritize those situations and locations where we require the absolutely best awareness we can acquire. This is true for cyberspace as well.

During the interwar period, fear of an attack from the air spurred several nations to bolster their nascent radio wave detection research program to improve their ability to detect an incoming air threat. The efficacy of building an elaborate detection network was put on display during the Battle of Britain. Their effort focused on detecting the threat as far away as possible, concentrating on the most probable avenue of attack. Similarly, in the early days of the Cold War, the United States used Ballistic Missile Early Warning sites to detect incoming Soviet intercontinental or submarine launched nuclear ballistic missiles. Physics determined the limited number of ways the Soviet Union could employ ballistic missiles against the US which in turn determined the number, type, and location of sensors we would have to build. Initially, only three sites were required to give adequate coverage against the threat. In both cases, geography, threat, and response time determined the type of detection required to defend against an attack.

Current efforts in cyberspace have focused heavily on Intrusion Detection Systems to identify when a network has been compromised. Unfortunately, while these sensors are necessary, they do not provide sufficient response time to react to a malicious attack. Ultimately, we would like to conduct deep packet inspection as far away from our network as possible, potentially in an isolated environment. Several technologies hold promise for conducting this type of early warning.

Still, at some point, network security will be breached. Just like there are no perfect radars or fences, there are no perfect intrusion detection systems that will detect 100% of the intrusions. For that reason, it is not sufficient to simply scan the borders of the network. Williams [4] suggests that the real damage of an intrusion is caused by the widespread and silent compromising of a system. "Organizations must monitor their systems for changes in connections between computers and servers, as well as patterns of mutations that seem to spread on their own."

Understanding the configuration of the network and the types of communication taking place on the network is a critical aspect of any defense.

Many "closed" networks operate under the assumption that whoever is on the network is authorized to be there. Situational awareness of the network becomes even more important for a "closed" network because of the sensitivity of the information on the network or the critical function it supports. For example, no matter how secure or "closed" the nuclear command and control network becomes, the possibility always remains that someone will get in. As networks proliferate and integrate, the ability to access a system undetected becomes easier. These types of critical networks require constant validation of all activities and processes occurring on every device within the network. Obviously, this is no small feat especially considering that these types of scans will compromise speed without a corresponding increase in computational capacity. Still, a mobile and active defense demands this level of situational awareness in order to respond to the intrusion threat.

D. Recover and Respond

Even the most sophisticated air defense systems are breached and facilities attacked. Stealth aircraft and advanced electronic warfare capabilities can be used to effectively blind the defense. In the same way that there are no perfect defenses against illegal border crossings or stealth aircraft, our networks will never be perfectly secured. With enough determination, an adversary will eventually defeat any defense, especially if it remains static. Like other types of defenses, we must anticipate the possibility that someone will eventually get into even our most secure networks. If done sufficiently, hardening and dispersal, will mitigate, if not defeat all together, the initial effects of most attacks. However, the adversary will adapt and the defense must react quickly. Once the network is breached, it becomes imperative to recover from the damage, find the threat, and respond.

Unlike any other domain, cyberspace can be redesigned. IP addresses can be changed, signals disrupted and new connections established, and routers, servers and switches taken off line while new ones are brought on. By reconfiguring the network and possibly moving data and functions to new segments or even new networks, cyberspace has the potential to be the most flexible and adaptive domain of warfare. For example, when components cannot be hardened against an electronic attack, alternate systems need to be available so that the network can be reconfigured or the data and functions rapidly moved to another network. Again, this is a capability that exists today. A recent study [10] demonstrated the ability to rapidly move functions across heterogeneous operation systems and platforms.

Maneuver warfare involves moving in relation to the adversary and conducting integrated movement across multiple domains [11]. Moving in relation to the adversary requires understanding the characteristics and physics of cyberspace as well as how a potential adversary uses cyberspace. At a tactical level, understanding movement in cyberspace means first understanding what is moving and how it is moved. Earlier, cyberspace was described as the place where data is created, stored, modified, and exchanged. What is moving is the data and how it is moved is through signals and electronics. Movement in cyberspace is accomplished by modifying either the signal (wireless or wired) or the software and hardware that manages the signal. If an adversary is targeting a particular signal frequency or internet protocol address to

disrupt or attack the data, then moving to a different type of signal or IP address would counter his attack tactically.

These concepts were recently summarized by a former chief scientist of the United States Air Force. In his final report [8], he concluded that a fundamental shift from protection to mission effectiveness would emphasize "technologies such as IP hopping, network polymorphism, massive virtualization and rapid network re-composition that can make cyber systems inherently resilient to intrusions entering through the network layer. These convert the currently static network layer into a highly dynamic one, in which the hypervisor mapping between the hardware and functional layers changes constantly in a pseudo-random way, perhaps hundred of times every second. A cyber adversary who finds vulnerabilities in the physical layer thus has virtually no time to use them for mapping the network before its topology has changed."

5. ANTICIPATING THE FUTURE

The current offensive advantage results from the ability to maneuver against a network combined with rapidly adaptive tools to attack networks and information. Current defense measures just simply cannot be prepared for the unknown and seemingly limitless ways to penetrate and attack a network. Increasingly, the most vulnerable networks are mobile. This past year, more smart phones were sold in the world than personal computers. This trend will continue across all types of networks; private, commercial, government and military. In fact, the US military is currently making plans to extend command and control infrastructures and increase access to information—including classified information—by distributing smartphones and tablet devices to individuals operating throughout the battlefield [12]. Even some of the newest satellites being tested are nothing more than smart phones placed in a box and launched into space [13]. Governments and private industry are exploring ways to use smartphones to build on orbit communications and sensor networks.

This expansion of network capability will provide a greater tactical and operational advantage, but also risks introducing even more vulnerability to the battle networks. The rapidly changing configuration of these highly mobile networks will be both a blessing and a curse for attackers and defenders. Implementing dispersal and hardening techniques however, could achieve a level of agility and protection for these types of networks that could result in an advantage for the defender rather than the attacker. In the same way that highly mobile, integrated air defense systems present a formidable challenge to attacks from the air, so too can data and communication networks achieve a similar level of capability.

One way to gain a position of advantage, particularly against a superior adversary, is to move to where he is weakest. This indirect approach applies force against an adversary's vulnerabilities. The social use of cyberspace represents another vulnerability but also an opportunity. Identifying who is part of an organization and their relation to others inside and outside the organization is essential to developing access to that organization. Prior to the explosion of information technology, this attack method was primarily conducted by covert agents who co-opted a member of organization to gain access and information. The ultimate covert action was to gain membership to the organization so that information could be accessed directly. Once

inside, they would have varying degrees of access to different types of information or even sensitive assets.

The prevalence of email and social networking sites make cyberspace an ideal medium to gain access to an organization. Unlike strangers, people who use information systems tend to trust the information they are presented. At the moment, it is relatively easy to deceive people in cyberspace and gain their trust. Further, the designed openness of social network sites introduces vulnerabilities to any organization's network. While social networking sites are typically riddled with malware and simply should not be accessed from mission critical systems, there is an opportunity to use these same vulnerabilities to expose potential adversaries. Configuring honey-pot networks using virtual machines, networks, and even cloud environments, may offer some ways to gain early warning of an attack and adversary techniques.

6. CONCLUSION

Sufficient capabilities exist today to counter the offensive advantage in cyberspace. What is lacking is an operational concept that can organize and integrate these capabilities into a posture that makes the defense more capable than the offense. This paper has introduced an integrated cyber defense strategy that increases network resiliency by dispersing and hardening the functions and data resident on the network. This includes taking advantage of network diversity to further complicate an attacker's problem. In addition, the defensive strategy relies on detecting the threat and adopting recovery procedures to respond the eventuality that a network will be breached. Together, these four characteristics of a integrated defensive system increase the strength of the defense and may even yield an advantage against the offense.

The uniquely dynamic nature of cyberspace, however, will ultimately shift the balance in favor of the defense. Highly mobile and hidden systems are extremely difficult to target. Despite the highest priority given to the mission, coalition forces were largely unsuccessful in eliminating the SCUD threat in Iraq during DESERT STORM. The Iraqi systems were easy to move and disguise which made them virtually impossible to find and target. Cyberspace has even more potential to be highly mobile allowing the defense to stay one step ahead of the offense and avoid an attack outright.

Networks continuously change and this change can be incorporated into a defensive strategy. The *tactical* advantage in cyberspace goes to those countries that can increase the speed and agility of their networks through more precise timing and increased processing power. However, the ability to rapidly establish, reconfigure, and distribute networks as well as the data and functions on the network will yield the *strategic* advantage in cyberspace.

REFERENCES:

[1] Joint Publication 1-02, Department of Defense Dictionary of Military and Associated Terms, September 2010.

[2] Julian. S. Corbett, *Some Principles of Maritime Strategy*, Annapolis: Naval Institute Press, 1988.

[3] Guilio Douhet, (1942) *The Command of the Air*, Washington, D.C.: Office of Air Force History, 1942.

[4] Paul Williams, Remarks given at FOSE information-technology exposition and conference in Washington, D.C., July 2011. Available: http://www.nationaldefensemagazine.org/blog/Lists/Posts/Post.aspx?ID=478

[5] Kamal Jabbour and Sarah Muccio, "The Science of Mission Assurance", *Journal of Strategic Security*, Vol IV, Issue 2, 2011, pp. 61-74.

[6] Keith Stouffer, Joe Falco, and Karen Scarfone, *Guide to Industrial Control Systems (ICS) Security*, National Institute of Standards and Technology, Special Publication 800-82, June 2011, p. 3-2.

[7] Ken Sorrels, Remarks given at FOSE information-technology exposition and conference in Washington, D.C., July 2011. Available: http://www.nationaldefensemagazine.org/blog/Lists/Posts/Post.aspx?ID=478

[8] "Technology Horizons: A Vision for Air Force Science & Technology During 2010-2030". [Online] Available: http://www.aviationweek.com/media/pdf/Check6/USAF_Technology_Horizons_report.pdf

[9] P. Williams and R. Boivie, "CPU Support for Secure Execution", Trust 2011, 4th International Conference on Trusted Computing, June 22-24, 2011, Pittsburgh PA.

[10] Hamed Okhravi, Adam Comella, Eric Robinson, Joshua Haines, "Creating a cyber moving target for critical infrastructure applications using platform diversity", *International Journal of Critical Infrastructure Protection*, 5, 2012, pp. 30-39. Available online: www.sciencedirect.com

[11] Richard D. Hooker, Jr., *Maneuver Warfare: An Anthology*. Novato, CA: Presidio Press, 1993.

[12] Steven Pugh, "A Top-Secret Smartphone Could Become Reality", *Signal*, November 2011, pp 15-16.

[13] "NASA ARC is Testing Cubesats on Balloons", SpaceRef Interactiv Inc. DBA SpaceRef Interational Group, [Online] http://www.spaceref.com/news/viewsr.rss.html?pid=36489

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

An Analysis For A Just Cyber Warfare

Mariarosaria Taddeo[1]
Department of Philosophy –
School of Humanities
University of Hertfordshire
Hatfield, UK
m.taddeo@herts.ac.uk

Abstract: This article focuses on the ethical analysis of cyber warfare, the warfare characterised by the deployment of information and communication technologies. It addresses the vacuum of ethical principles surrounding this phenomenon by providing an ethical framework for the definition of such principles. The article is divided in three parts. The first one considers cyber warfare in relation to the so-called information revolution and provides a conceptual analysis of this kind of warfare. The second part focuses on the ethical problems posed by cyber warfare and describes the issues that arise when Just War Theory is endorsed to address them. The final part introduces Information Ethics as a suitable ethical framework for the analysis of cyber warfare, and argues that the vacuum of ethical principles for this kind warfare is overcome when Just War Theory and Information Ethics are merged together.

Keywords: *cyber warfare, information ethics, Just War Theory*

1. INTRODUCTION

During the past two decades, information and communication technologies (ICTs) proved to be a useful and convenient for war waging, so much so that they have been deployed in most of the conflicts since the second Iraq's war.[2] The military deployment of ICTs has radically changed the way wars are waged nowadays. It has actually determined the latest revolution in military affairs, making the cyber space the fifth domain of war, along with land, sea, air and space.

The informational turn in military affairs is not of exclusive concern of the militaries; it also concerns ethicists and policymakers. For existing ethical theories of war and national and international regulations struggle to address the novelties of this phenomenon. This article is devoted to develop an ethical analysis of cyber warfare (CW), with the twofold goal of overcoming the theoretical vacuum surrounding this phenomenon and of providing the grounding for an ethical regulation for CW.

The proposed analysis rests on the investigation of CW proposed in (Taddeo 2012), which highlights the informational nature of this phenomenon as well as its relation to the so-called

1 This paper is part of the research supported by the Marie Curie Intra-European Fellowships.
2 *See* http://www.economist.com/node/16478792.

Information Revolution. In this paper it will be argued that Just War Theory (JWT) is a necessary but not sufficient instrument for the ethical analysis of CW. It will be maintained that analysing CW through the lenses of JWT allows for unveiling the fundamental ethical issues that this phenomenon brings to the fore, but that attempting to address these issues solely on the basis of JWT will leave them unsolved.

The thesis will be advanced that the problems encountered when addressing CW through JWT are overcome when the latter is merged with Information Ethics (Floridi 2008). This is a macro-ethical theory developed to take into account the features and the ethical implications of *informational phenomena*, like internet neutrality (Turilli et al. Forthcoming), online trust (Turilli et al. 2010), peer-to-peer (Taddeo and Vaccaro 2011) and CW. The goal is to develop an ethical analysis of CW able to take into account both its peculiarities and its novelty, while at the same time be consistent with the mainstream ethical analysis of warfare.

Having delineated the path of the analysis proposed in this article, we shall now begin by considering in more details the nature of CW.

2. CYBER WARFARE

For the purpose of this article CW is defined as follows:

> "**[Cyber] Warfare** is [the warfare grounded on certain] uses of ICTs within an offensive or defensive military strategy endorsed by a state and aiming at the immediate disruption or control of the enemy's resources, and which is waged within the informational environment, with agents and targets ranging both on the physical and non-physical domains and whose level of violence may vary upon circumstances", (Taddeo 2012, 114).

This definition highlights two aspects of CW, its *informational nature* and its *transversality*[3]. The informational nature of CW is a consequence of the fact that this kind of warfare rests on the military deployment of technological artefacts devoted to elaborate, manage and communicate data and information. With this respect CW shows to be related to the so-called Information Revolution.

The Information Revolution is a multi-faced phenomenon. It rests on the development and the capillary dissemination of the use of ICTs, which have a wide impact on several of our daily practises, from working, to interacting with other human beings, to driving around and planning holidays. The dissemination of ICTs has important philosophical implications (Floridi 2010), for the Information Revolution changes fundamentally the way reality is perceived and understood.

Information Revolution determines a shift, which brings the *non-physical domain* to the fore and makes it as important and valuable as the physical one. CW is one of the most compelling instances of such a shift, it shows that there is a new environment, where physical and non-physical entities coexist and are equally valuable, and in which states have to prove their

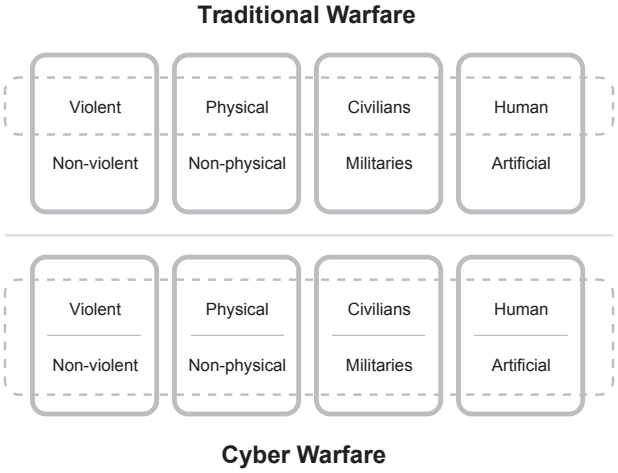
3 'Transversality' is used in this article to indicate that CW cuts across any qualifying couple such as 'violent-non violent', 'civil-military', 'human agents-artificial agents'. This aspect is quite different from traditional warfare, which is violent, conducted by militaries and mainly by human agents.

authority and new modes of warfare are being developed specifically to be deployed in such a new environment (Taddeo 2012).[4]

The shift toward the non-physical domain provides the ground for the transversality of CW. This is a complex aspect, and can be better grasped when CW is compared with traditional form of warfare. Traditional war is understood as the use of a state's *violence* through the state *military* forces to determine the conditions of governance over a determined territory (Gelven 1994). It is a necessarily violent phenomenon, which implies the sacrifice of human lives and the damage of both military and civilian infrastructures. The problem to be faced when waging traditional warfare is how to reduce to the minimum such damages while ensuring to overpower the enemy.

CW shows to be different from traditional warfare, as it is not a necessarily violent and destructive phenomenon (Arquilla 1999). CW may involve a computer virus able to disrupt or deny access to the enemy's database, and in so doing cause a severe damage to the enemy without exerting *physical* force or violence. In the same way, CW does not necessarily involve human beings. An action of war in this context can be conducted by a computer virus, targeting other artificial agents or informational infrastructures, like a database or a website (see Figure 1). Nevertheless, CW is to be feared as much as traditional warfare, for it is transversal with respect to the level of violence and may escalate from non-violent to more violent forms. Consider, for example, the consequences of a cyber attack targeting a military aerial control system causing aircraft to crash (Waltz 1998). As remarked above, the transversality of CW with respect to the levels of violence, the nature of the agents and the waging domain is the key feature of this phenomenon, the aspect that differentiates it the most from traditional warfare, and also the feature that engenders the ethical problems posed by CW.

FIGURE 1: CW COMPARED TO TRADITIONAL WARFARE IN RESPECT TO THE COUPLES 'VIOLENT AND NON-VIOLENT', 'CIVILIANS-MILITARIES', 'HUMAN AND ARTIFICIAL AGENTS', 'PHYSICAL AND NON-PHYSICAL'. THESE COUPLES ARE EMBLEMATIC OF THE KIND OF WAR WHICH IS WAGED AS THEY IDENTIFY



4 The USA only spent $400 million in developing technologies for cyber conflicts: see http://www.wired.com/dangerroom/2010/05/cyberwar-cassandras-get-400-million-in-conflict-cash/. The UK devoted £650 million to the same purpose: see http://www.theinquirer.net/inquirer/news/1896098/british-military-spend-gbp650-million-cyber-warfare.

Transversality makes CW extremely appealing from both an ethical and political perspectives (Arquilla and Ronfeldt 1997). At first glance, CW seems to avoid bloodshed and human commitment and therefore it liberates political authorities of the burden of justifying military actions to the public opinion. A more attentive analysis unveils that CW should be feared as much as traditional warfare as it can lead to highly violent and destructive consequences, which could be dangerous for both the military forces and civil society.

For this reason, declaring and waging CW require a strict ethical regulation to guarantee its fairness. An analysis of CW unveiling the ethical issues that it engenders and pointing at the direction for their solution is a necessary step toward the achievement of such goal.

3. JUST WAR THEORY AND CYBER WARFARE

JWT refers to war as to a violent and sanguinary phenomenon, declared by states and their official leaders and waged by military forces. Such a scenario is quite different from the one determined by CW, the difference between the two forms of warfare is the origin of the problems arising when the principles of JWT are applied to CW. In this respect, there are three issues that deserve attention; they follow from the application of the principles of 'war as last resort', of 'more good than harm', and of 'non-combatants immunity' to CW.

As highlighted in (Taddeo 2012), the application of the principle of 'war as last resort' is shaken when CW is taken in consideration, because in this case war may be bloodless and may not involve physical violence at all. In these circumstances, the use of the principle of war as last resort becomes less immediate.

Imagine, for example, the case of tense relations between two states and that the tension could be resolved if one of the states decides to launch a cyber attack on the other state's informational infrastructure. The attack would be bloodless as it would affect only the informational grid of the other state and there would be no casualties. The attack could also lead to resolution of the tension and avert the possibility of a traditional war in the foreseeable future. Nevertheless, according to JWT, the attack would be an act of war, and as such it is forbidden as a first strike move. The impasse is quite dramatic, for if the state decides not to launch the cyber attack it will be probably forced to engage in a sanguinary war in the future, but if the state authorises the cyber attack it will breach the principle of war as last resort and commit an unethical action, which could probably be sanctioned by international regulations.

This example is emblematic of the problems encountered in the attempt to establish ethical guidelines for CW. In this case, the main problem is due to the transversality of the modes of combat, which make it difficult to define unequivocal ethical guidelines. In the light of the principle of last resort, soft and non-violent cases of CW can be approved as means for avoiding traditional war (Perry 1995), as they can be considered a viable alternative to bloodshed. At the same time, even the soft cases of CW have a disruptive purpose – disrupting the enemy's (informational) resources (Floridi 2008) –, which needs to be taken into consideration by any analysis aiming at providing ethical guidelines for CW. Even when the disruption of the enemy's informational infrastructure is not achieved through violent and sanguinary means.[5]

5 For a more in depth analysis of the non-violent cases of CW and their assessment as acts of war or of espionage *see* (Arquilla 1998) and (Taddeo 2012).

The second problem to be considered concerns the principle of 'more good than harm'. According to such a principle, a state is justified in declaring war only when the goods are proportional to the evils. This balance is easily assessed in case of traditional warfare, where the evils are mainly considered in terms of the casualties and physical damages. The equilibrium between the goods and the evils becomes more problematic to determine when CW is taken under consideration.

CW is likely to cause none or very little casualties, and as it targets informational infrastructures it is unlikely to cause the destruction of physical objects, like buildings for example. Although it is possible for CW to turn in a violent warfare, in the most of the cases it does not determine physical damages, nonetheless CW may result in unethical actions. If the only criteria for the assessment of the harm in warfare scenario remain the consideration of the physical damages caused by war, then an unwelcome consequence follows. For all the non-violent cases of CW comply by default to this principle. Therefore, destroying a digital database or erasing a digital archive containing important historical records of a nation are all deemed to be ethical actions as they do not constitute *per se* a physical damage.

In the case of this principle, it is not the prescription that the goods should be greater than the harm in order to justify the decision to wage a war to be shaken. It is rather the set of criteria to assess the good and the harm, which show to be inadequate when considering CW.

The last problem concerns the principle of 'discrimination and non-combatant immunity'. Also this principle refers to a classic war scenario and aims at reducing the bloodshed and prohibits any form of violence against non-combatants, like civilians. Its correctness is not questionable yet its application is quite difficult in the context of CW.

In classic warfare, the distinction between combatants and non-combatants reflects the distinction between military and civil society. Even if the diffusion of terrorism and guerrilla warfare during the 20th century weakened the association between non-combatants and civilians, in the case of CW such association becomes even feebler, due to the blurring between civil society and military organisations (Schmitt 1999; Shulman 1999).

As noted in (Taddeo 2012), the blurring leads to the involvement of civilians in war actions and poses two issues. The first one concerns the discrimination itself: in the CW scenario it is difficult to distinguish combatants from non-combatants, wearing a uniform is no longer a sufficient criterion to identify someone's status. Civilians may take part in a combat action from the comfort of their homes, while carrying on with their civilian life and hiding their status as cyber warriors.

The second issue concerns the effects of this difficulty in distinguishing combatants from non-combatants and unveils an ethical conundrum. If combatants can easily hide themselves among the civilian population, then states may be justified in endorsing high levels of surveillance over the entire population, thereby breaching individual rights, like privacy and anonymity, in order to identify the combatants and guarantee the security of the entire community. For the sake of these goals, public authorities could also be justified in persecuting certain sections of the civilian population, which are profiled and deemed to be potentially dangerous for the

community. Therefore, on the one side respecting the principle of discrimination may lead to the violation of individual rights. On the other side, waiving the principle of discrimination leads to bloodshed and dissemination of violence over the entire civil population, because the policy could be endorsed to target everyone or everything a soldier encounters in her way, as being potentially involved in the conflict.

It would be misleading to consider the problems described in this section as reasons to disregard JWT when analysing CW. The ideal of just warfare provided by JWT and its principles remain valid even when considering this new kind of warfare. Yet, the analysis proposed in this section points to a more fundamental problem, namely the need to provide an ethical framework for the regulation of CW able to address the novelty of this phenomenon. In the next section, Information Ethics will be introduced as the suitable ethical framework for this purpose.

4. INFORMATION ETHICS

Information Ethics is concerned with the ethical issues in which information is involved as a resource, as a product, and as a target (Floridi 2008a). It proposes a twofold approach: (i) considering the whole information-cycle, from creation, to communication and storage, and (ii) analysing *informationally* all entities involved in a moral scenario. The moral agents and their actions are considered as part of the informational environment to which they belong as informational entities themselves (Taddeo and Vaccaro 2011).

In this framework, two concepts are of pivotal relevance: Infosphere and informational ontology. As remarked in (Taddeo and Vaccaro 2011), the Infosphere is the totality of what exists. The Infosphere includes agents and objects, relations and processes, as well as the space within which they act. It is not to be confused with cyberspace, as it includes online as well as offline and analogue domains. Infosphere comprises e-books and trees, online websites and rocks, movies in digital format and the paintings on canvas.

The Infosphere is the environment in which animate and inanimate, digital and analogue informational objects are morally evaluated. Information Ethics endorses a universal approach, according to which all exiting things, i.e., not only human beings and living things, but also artefacts and digital artefacts enjoy some minimal and overridable moral rights (Taddeo and Vaccaro 2011).

This universal perspective is grounded in an ontocentric principle, according to which all entities, understood as informational objects, have the fundamental rights to exist and flourish. In Floridi's words: '[...], any form of reality (any instance of information/being), simply by the fact of being what it is, enjoys a minimal, initial, overridable, equal right to exist (be left alone) and develop (not to be interfered) in a way which benefits its nature' (Floridi 2007b).

In such a universal context, the morality of a given action is assessed with respect to the effects that it will have on the patients, i.e., the recipients of the action, and ultimately on the Infosphere. This is referred to as the patient-oriented perspective of Information Ethics, according to which, we can decide whether an action is evil only on the basis of a clear understanding of its effects on interacting patients.

In a nutshell, Information Ethics is an environmental ethics, which endorses an ontocentric and patient-oriented approach, and in which the morality of a course of action is evaluated on the basis of its effects on informational entities and ultimately on the Infosphere. (Floridi 2008a).

Within this framework, Information Ethics provides four moral principles that ought to be respected in order to preserve the well-being and continued flourishing of the Infosphere and its inhabitants:

0. Entropy ought not to be caused in the Infosphere (null law);
1. Entropy ought to be prevented in the Infosphere;
2. Entropy ought to be removed from the Infosphere;
3. The flourishing of informational entities as well as the whole Infosphere ought to be promoted by preserving, cultivating, enhancing and enriching their properties.

The concept of *entropy* adopted in the four laws indicates the result of any form of 'destruction, corruption, pollution, depletion (marked reduction in quantity, content, quality, or value) or unjustified closure of the Infosphere' (Floridi 2001). Informational entropy is the evil, which should be avoided in the Infosphere and should be understood as a metaphysical concept, and it is not related to the concept of physical entropy or the use of entropy made in Shannon's information theory.

Now that the ethical principles and the approach endorsed by Information Ethics have been described, we can focus on its application to CW.

5. JUST CYBER WARFARE

Following the ontocentric approach, all (informational) entities enjoy some minimal rights to exist and flourish in the Infosphere. As such all entities, would they be leaving things or non-living things, physical or virtual, deserve some minimal respect. When applied to CW, this principle allows for considering as moral patients all the entities that may be affected by an action of war within CW. A human being, who suffers the consequences of a cyber attack and an informational infrastructure that is disrupted by a cyber attack are both to be consider the receiver of the moral action. The morality of that action will be assessed on the basis on its effect on their rights to exist and flourish.[6]

The first question when considering the conditions for a just CW concerns the rights of the informational entities, namely what and whose rights should be preserved. The answer to this question follows from the rationale of Information Ethics. Information Ethics states that an entity looses it rights to exist and flourish when it comes into conflict with the rights of other entities or with the well-being of the Infosphere. Therefore, any entity that causes entropy in the Infosphere loses its informational rights as it conflicts with the well-being of the other entities and ultimately of the Infosphere. It is a moral duty of the other inhabitants of the Infosphere to

6 While assuming that all entities share some initial rights to exist and flourish, Information Ethics does not claim that there is no hierarchy among the entities. It specifies that the rights are overridable and hence that an entity ceases to hold the rights to exist and flourish, should it contravene the well-being of other entities or of the Infosphere. Furthermore, according to Information Ethics, the position in the hierarchy of an entity depends on its contribution to the flourishing of the Infosphere. For a more in depth analysis of the criteria to override the entities initial rights *see* (Floridi 2008).

remove such a malicious entity from the Infosphere, as it is a cause of entropy, or to impede it to perpetrate more evil.

This lays the ground for the first principle for just CW. The principle prescribes the condition under which the choice to resort to CW is morally justified:

I. CW ought to be waged only against those entities that endanger or disrupt the well-being of the Infosphere.

Two more principles regulate just CW, they are:

II. CW ought to be waged to preserve the well-being of the Infosphere.
III. CW ought not to be waged to promote the well-being of the Infosphere.

The second principle limits the task of CW to restore the *status quo* in the Infosphere before the malicious entity began increasing the entropy in it. According to the second principle, CW should act only when some evil has been or is about to be perpetrated with the goal of stopping it. CW ought to be endorsed as an *active* measure in response to the increasing of the evil and not as *proactive* measure to foster the flourishing of the Infosphere. This is explicitly forbidden by the third principle, which prescribes that the promoting of the well-being of the Infosphere does not pertain to the scope of a just CW.

The time has come to consider how JWT can be applied to the case for CW without leading to the conundrums described in section 3.

6. THREE PRINCIPLES FOR A JUST CYBER WARFARE

The application of the principle of 'last resort' provides the first instance of how JWT and Information Ethics are merged. The principle takes into account traditional (violent) forms of warfare, and it is coupled with the principle of 'right cause', which justifies the resort to war only in case of 'self-defence'. As much as rightful this approach is when referred to traditional (violent) form of warfare, it shows to be inadequate when CW is taken under consideration. The impasse is overcome when considering the principles for just CW.

The first principle prescribes that any entity that endangers or disrupts the well-being of the Infosphere loses its basic rights and becomes a licit target. Therefore, a state can rightly endorse CW as an early move against a malicious entity. The choice to resort to CW is furthermore justified if it allows a state to avoid the possibility of a traditional warfare, as this one would determine casualties and destructions in the Infosphere, and as such it is deemed to be a greater evil than CW.

A caveat must be stressed in this case; the waging of CW must comply with the principles of 'proportionality' and 'more good than harm'. In waging CW, the means endorsed to win the enemy must be sufficient to stop the malicious entity, yet they ought not to generate more entropy than the one a state is aiming to remove from the Infosphere. This leads us to consider in more detail the principle of more good than harm.

The application of this principle is of paramount importance for the waging of a just warfare, would it be a traditional or an informational one. As noted in section 3, the issues concerning CW are due to the definition of the criteria for the assessment of the 'good' and the 'harm' that warfare may cause. Traditionally, they are defined with respect to the collateral damage, casualties, and damages to the physical infrastructures of both the parts involved in the war. Such criteria do not take in consideration the harm that CW may cause.

In the case of CW, the damage to non-physical entities needs to be considered as well as the damage to the physical ones. More precisely, the assessment of the good and the harm should be determined considering the general condition of the Infosphere 'before and after' waging the war. A just war never determines greater entropy (evil) than the one that it intended to remove from the Infosphere in the first place. Once considered in this perspective, the principle of more good than harm acts as corollary of the second principle for just CW. It ensures that a just CW is waged to restore the *status quo* and it never increases the level of entropy in the Infosphere.

The assessment of the entropy in the Infosphere allows also for reconsidering the application of the principle of non-combatants immunity to CW. Two problems accompany the application of this principle, the consequences of its endorsement on the individuals' rights of privacy and anonymity, and the very distinction between combatants and non-combatants. The rest of this section will focus only on the latter issue; the former does not pertain to the scope of this paper and as such will not be considered here.[7]

The distinction between combatants and non-combatants promoted by this principle rests on the distinction between militaries and civilians that is inherited from traditional warfare. As we have seen, CW is transversal with respect to the social status of the combatants, for it does not require military skills to be waged. This makes problematic the application of the principle, which nevertheless has to be maintained as it prescribes the distinction between enemies and 'innocents'.

Help in applying this principle to CW comes from the first principle for just CW, which allows for overcoming the distinction between militaries and civilians, and for substituting it with the distinction between licit targets and non-licit ones, the former being the malicious entities that endangered or disrupted the well-being of the Infosphere.

The time has arrived to pull together the threads of the analysis proposed in this article.

7. CONCLUSION

This article rests on the conceptual analysis of CW provided in section 2. Such analysis stresses the novelty of this phenomenon, its relation with the Information Revolution and argues that transversality is its main feature. Transversality is deemed to be the characteristics of CW that differentiates it the most from traditional warfare and also the one from which all the ethical issues posed by CW originate.

It has been argued that, given the radical novelty posed by CW, the ethical analysis of this

7 For an in depth analysis of this issue *see* (Taddeo 2012).

phenomenon and the definition of the ethical principles for a just CW cannot rest solely on JWT. For such a theory does not provide 'the right sieve' for the work to do. JWT does not take into account the main features of CW, namely the transversality of the levels of violence, of the domain (physical and non-physical) in which it is waged, and finally the transversality of the nature and social status of agents who may be involved in this warfare. Yet, the article maintains that it would be mistaken to reject JWT altogether when addressing CW.

It is rather argued that the ideal of just warfare and the principles prescribed by JWT are still valid when referred to CW, and that they can be endorsed to regulate this new form of warfare if they are combined with a macro-ethical framework able to take into account the peculiarities of this phenomenon.

Information Ethics has been introduced as a suitable ethical framework for CW. This is a macro-ethics, which endorses an ontocentric, patient-oriented and ecological approach and is devoted to address the ethical problems posed by informational phenomena. In particular, the ecological facet of Information Ethics shows to be extremely relevant for the purpose of the analysis proposed in this article, as by posing the well-being of the Infosphere as the ultimate good and the creation of entropy in the Infosphere as the moral evil, it provides the criteria for the ethical assessment of the implications of CW.

Three principles for just CW, encompassing both the rationale of JWT and of Information Ethics, have been provided. Such principles constitute the grounding for the development of more detailed ethical guidelines for CW that is for the next step of this research.

REFERENCES

Arquilla, John. 1998. "Can information warfare ever be just?" *Ethics and Information Technology* 1(3): 203-212.

Arquilla, John. 1999. "Ethics and information warfare." *In Strategic appraisal: the changing role of information in warfare*, edited by Z. Khalilzad, J. White, and A. Marsall, 379-401. Santa Monica, USA: Rand Corporation.

Arquilla, John, and Ronfeldt, David. 1997. *In Athena's Camp: Preparing for Conflict in the Information Age*. Santa Monica, CA: RAND Corporation.

Floridi, Luciano. 2008. "Information Ethics, its Nature and Scope." *Information Technology and Moral Philosophy Vol. 40-65*. Cambridge: Cambridge University Press.

Floridi, Luciano. 2010. The Digital Revolution as The Fourth Revolution. *Invited contribution to the BBC online program Digital Revolution*.

Gelven, Michael. 1994. *War and Existence*. Philadelphia, PA: Pennsylvania State University Press.

Perry, David L. 1995. "Repugnant Philosophy: Ethics, Espionage, and Covert Action." *Journal of Conflict Studies, Spring*.

Schmitt, Michael N. 1999. "The Principle of Discrimination in 21st Century Warfare." *Yale Humna Right and Development Law Journal* 2:143-160.

Shulman, Mark R. 1999. "Discrimination in the Laws of Information Warfare." *Pace Law Faculty Publications* 37:939-968.

Taddeo, Mariarosaria. 2012. "Information Warfare: a Philosophical Perspective." *Philosophy and Technology*, 25(1): 105-120.

Taddeo, Mariarosaria, and Vaccaro, Antonino. 2011. "Analyzing peer-to-peer technology using information ethics." *The Information Society* 27(2):105 - 112.

Turilli, Matteo, Vaccaro, Antonino, and Taddeo, Mariarosaria. (2010). The Case of on-line Trust. *Knoweldge, Technology and Policy* 23(3-4):333-345.

Turilli, Matteo, Vaccaro, Antonino, and Taddeo, Mariarosaria. (Forthcoming). *Internet Neutrality: Ethical Issues in the Internet Environment*.

Waltz, Edward L. 1998. *Information Warfare Principles and Operations*. Norwood, USA: Publisher Artech House, Inc.

Chapter 4
Cyber Conflict – Actors

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Socially Engineered Commoners as Cyber Warriors – Estonian Future or Present?

Birgy Lorenz
Institute of Informatics
Tallinn University
Tallinn, Estonia
birgy.lorenz@tlu.ee

Kaido Kikkas
Institute of Informatics
Tallinn University
Tallinn, Estonia
kaido.kikkas@tlu.ee

Abstract: The goal of our paper is to find out the readiness in Estonia to raise awareness of cyber security-related social engineering, especially among common people. We suggest that the awareness and understanding of online social engineering can raise the Estonian defence potential to a new level. Future cyber attacks may complement server attacks with human engineering and spreading misinformation in order to create incentives for treason or mutiny against the decisions of the state. Social engineering of information and people is one way to wage modern information wars.
Estonia is probably closer than anyone else to a functioning e-society, so it is important to build it up as safely and trustworthy as possible, inform people about potential downsides and suggest solutions for them. Due to widespread adoption of various e-solutions, the Estonian situation of e-safety and awareness could be considered adequate, but it can also turn out to be a weakness. Trust in the e-government, e-police, e-tax office etc. can lead to complete trust in e-channels as a whole, in turn creating extensive dependence on them.
We have conducted a study involving schoolchildren and ICT students, as well as members of the Estonian Defence League Cyber Defence Unit (EDLCDU). The findings suggest a way to carry out related training programmes or campaigns. The recommendations are useful for coordinating the efforts of the four Ministries involved, addressing the crossroads of technical cyber security, social interaction, communication and education.

Keywords: *: cyber security; social engineering; education policy*

1. INTRODUCTION

Our hypothesis is that in the cyber war situation, when information is scarce and the circumstances are hard to understand, ordinary citizens (lay people) can turn against their government and critical services (e.g. public transport and infrastructure), carrying the conflict over from the cyberspace into actual space – but this kind of process can be largely prevented

by proper policies as well as education.

Moreover, it is possible to turn the liability into an asset, using adequately trained and motivated lay people as a kind of "cyber militia" to complement the efforts of "regular forces" or cyber defence specialists. In Estonia, this has already been partially achieved in the form of the Estonian Defence League Cyber Defence Unit (EDLCDU) which was founded after the 2007 cyber-attacks after the Bronze Soldier riots [1].

We expand the term "cyber warfare" from politically-motivated attacks on systems (to conduct sabotage or espionage) to large-scale manipulation of information (media, government, hackers) and potential crowd control in this situation. We believe that understanding chain reactions in this area provides valuable information to governments acting in crises, SCADA (supervisory control and data acquisition) units and different Ministries whose responsibility should be raising awareness. Thus our goal is to find out the public stance on the implementation of cyber war-related training in elementary, secondary and higher education.

2. BACKGROUND

A. Similar Studies

The digital landscape has developed from the initial technological phenomenon into a complex and increasingly social one (social engineering, new applications and interpersonal trust), including new types of devices, applications and end-systems (e.g. iPads, Facebook, e-Banking, iPlayer, etc.), as well as network and infrastructure vulnerabilities (e.g. network attacks, failures and misconfiguration) [2]. Modern cyber-attacks have more to do with manipulating humans than ever [3].

The Information Technology education Model Curriculum still discusses whether cyber security should be a part of the programme or if it is something that is unique in the field [4]. For designers of security systems it is important to understand how users evaluate and make decisions regarding security [5]. Ordinary people do not think about the risks at home, why should they be better at work?

The Internet habits of adolescents have changed. Social behaviour, belonging and being a part of something is more important than ever [6]. Social engineering is considered a low-cost and effective form of attack because of the lack of awareness in this matter [7]. Some feel that social networking raises also the risk of automated social engineering, hijacking and phishing [8]. The weakest link is human behaviour [9], which has been largely missing from systematic analysis compared with other aspects of cyberspace [1].

B. The Changing World

Robert Theobald, an American futurist, has used the term "mind-quake" to denote a situation where an old dominant way of thinking is overridden by undeniable new understandings [11]. A good illustration in a recent context is provided by Rick Falkvinge (a Swedish IT entrepreneur), recalling the once well-established business of selling ice for cooling foodstuff during summertime and its subsequent fall after electric refrigerators became available [12].

A similar change of mentality occurred when Gutenberg introduced printing to Europe. It allowed people to spread knowledge and boost education. The Internet has done the same on an even larger scale – people are spreading the word and sharing materials etc. The Internet's main point is to spread, not to restrict, data sharing [13].

But as some kinds of data should not be accessible for everyone, it also leads to a possible way to manipulate people – one can feed them false information or sow distrust towards leaders or a currency etc. [14]. When this is done by governments to their own subjects, it is considered an internal affair (e.g. in Belarus or Syria). But when other governments or politically motivated groups intervene with other countries' politics by influencing the residents (e.g. Nashi [15]), it is much more likely to be considered as a psychological/cyber-attack or a case of social engineering – and if it happens repeatedly it can be considered to be a cyber-war.

Finally, there seems to be a growing need for two different terms for the current "cyber warfare" – one to denote cyber-attacks during military operations leading to a military objective (e.g. blinding enemy drones or radars) and another to reflect the emerging tendencies that also relate to cyber-attacks but use humans as a primary asset (but can produce similarly effective outcomes to straightforward military activity).

C. The Human Factor in Cyberwar

A definition of cyber war [16] states that:

- there should be consequences in real life;
- it is detectable afterwards;
- there are no persistent solutions that we can rely on;
- there are no limits to the physical distance;
- both sides (attacker and defender) have same rights and use same tools;
- whoever controls the opponent's resources controls the opponent.

In most of the current policies, the main priorities to be protected in the occurrence of a cyber-attack are either data or hardware – to detect the intrusion and regain control of, clean and patch the systems [17]. On the one hand, it is understandable; the government's main concern is to keep up critical services like finance, sustenance, medical assistance, transport, water and electricity, ICT services and public administration [18].

On the other hand, the role of patching the "human factor" has been seriously neglected. For example, playing with human thoughts and behaviour can incite devaluation panic (e.g. attempts to influence the Russian-speaking population in Estonia before the adoption of the euro in 2007) [19] or massive unrest (e.g. the Bronze Soldier affair in Estonia in 2007 [20] or similar events in Denmark in 2008 [21], the UK [22] or France [23]), as well as influence financial markets or incite large-scale protests (Middle-East, Egypt [24]). A recent example is the influx of malware using social engineering techniques after disasters like the Japan earthquake in 2011 [25].

D. Psychological and Sociological Factors

There are three ideas that contribute to the possibility of mass manipulation: we are all connected, tend to overreact and "with the right weather conditions, all Hell can break loose". So, the following points apply:

- In the age of social networking, "who knows who" has gained major importance. The concept of "six degrees of separation" refers to the idea that everyone is on average approximately six steps away. Nowadays we see it happening in real social networks [26]. For example, I know my country's political leader, who knows the President of the US, who knows everybody. So the steps can be even less numerous [27];
- Positive feedback loops are well known to describe the dynamics of change in biological evolution. Today, the same effects are seen on the net: a small disturbance launches several opinions from others and the result will be greatly amplified [28,29]. Sometimes it will not last long, but in cases of larger public interest [30,31] it can end up gathering to dance like Michael Jackson [32]. Essentially the same process worked for the Bronze Soldier riots or more currently the Arabian Spring, Occupy Wall Street [33] or the ongoing protests against ACTA [34];
- Nowadays the police consider weather conditions as one of the key elements to influence the risk of massive unrest. The threat is lower when it is too cold or too warm [35,36].

Kalev Leetaru, a computer scientist, has claimed that "pooling together the global tone of all news mentions of a country over time appears to accurately forecast its near–term stability, including predicting the revolutions in Egypt, Tunisia, and Libya, conflict in Serbia, and the stability of Saudi Arabia [37]".

Putting these three ideas together, we get an explosive mix. For example, Mr. Smith hears from a friend's friend that the Euro will be devalued in a few days. How would he act? When that information is fed to the public, how can the government be quicker and more reliable than the biased (as seen by many people) mass media or other Web 2.0 tools and social networks? Also (and perhaps most importantly), who is the enemy to blame?

E. Changes in Cyber-Warfare

Some would argue that, in recent years, the rapid development of technology has outrun the capability of governments to keep pace with it, while others would assume that Moore's Law is still valid [38]. The intervals between technology renewals have shortened – it is common to have a new smartphone every year and a new computer every two years [39]. Common people possess adequate (and rapidly evolving) computing power which cannot be sufficiently neutralised in cases of misuse or hijack. To make things worse, legislation lags far behind the situation and the processing queue of online crime-related cases is long and increasing [40-42].

All this makes detecting cyber incidents and forensics difficult; attackers evolve more quickly than defence [43,44]. Yet, manipulating people's mindsets can be even more devastating than getting unrestricted access to a service or server.

In general, the countries which have recently experienced some form of cyber conflict (e.g. Estonia or Russia) also tend to be more conscious in terms of related policies. The countries which have a strict "command line" (formal or informal) in either the government (e.g. Belarus or China) or a parallel structure (e.g. trade unions, CDL, diaspora, organised crime etc.) or the ones without strong dependence on ICT (e.g. some developing countries) are generally more resistant to cyber war [45].

F. The Situation in Estonia

"The Estonian way in cyber war issues is above all defence-oriented. Historically, Nordic Finno-Ugric tribes traditionally lived in peace with nature and neighbours. It's a lifestyle. At the same time, awareness against threats (cold climate, predators etc) has always been a normal part of life. This way, passively defending itself against threats, adjusts itself well to the Internet threats. It is important to notice and keep in mind that when Finno-Ugric people say "defence", it really is defence only – defending their lifestyle – and it is not including any deep hidden aggression or hidden agenda," says Anto Veldre (a cyber security specialist from Estonia).

In Estonia, raising the layperson's awareness in e-safety belongs to the domain of the Ministries of Education and Social Affairs, as well as Economy and Communication. However, it has recently also caught the attention of the Estonian Ministry of Defence, leading to the formation of the EDLCDU. Adults are usually trained by universities, voluntary trainers, media, workplaces and also schools (via children). It is still easier to train students and teachers by adapting school curricula – presuming that children will grow up to knowledgeable citizens or maybe even influence their parents and grandparents.

There is a Masters programme at the Tallinn University of Technology focusing on technical aspects of cyber security, yet no one is currently working on the lower stages of education [46]. There are some efforts supported by the EU (e.g. the InSafe programme) [47] as well the business sector programme "Be Included!" by the Look@World Foundation, that trained approximately 100,000 elderly people to use the national ID card and hence increase their security online [48]). There is the Cyber-Defence League which harbours both IT security experts and military personnel, focusing on educating its members. However, the National Defence curriculum sponsored by the Estonian Ministry of Defence does not currently focus on the layperson's cyber security awareness.

3. METHODS

We used triangulation with an interview and a survey to get better understanding of human behaviour and people's attitudes towards cyber war issues in their life:

- stage I – an interview with three cyber security experts, where we also got input to the stage II questionnaire from;
- stage II – the focus group study. We used the open source web application called Limesurvey. We collected 98 responses over two weeks;
- stage III – the results were analysed by five experts (two representing education, two ICT, one legislation), using the group analysis method.

Stage II focused on: students (42), 27 from secondary school, others from university (eight ICT related); ICT experts (25), six were EDLCDU-related; and other adults (31), 14 in education and two were EDLCDU-related. We used a survey with 37 questions divided into six sub-categories: general, government and cyberspace, cyber war means, ethics, incident response and background information. The data was collected using the Likert scale [49] and Q Methodology [50] rankings, plus open questions that were mainly used for clarification.

4. RESULTS

For background information we collected the participants' thoughts about understanding information from cyberspace. We found out that the most trusted channels they get information from are still traditional (TV, radio, newspaper – 58%). ICT specialists tend to also trust forums (12%), as do EDLCDU members (13%). EDLCDU members trust mailing lists (17%). Secondary school students (16%) are more open to online news, as are ICT specialists (11%). All participants trust academic texts (23%) more than any other media, including European (19%), local (16%), Northern (13%) and World (12%) news. It is interesting to note that even an informal chat with a friend (8%) is more trusted than official news channels from Russia (1%).

To understand what is perceived as cyber war and what is not, we described different scenarios and asked whether they are seen as such. The respondents seemed to consider it as something strictly related to online attacks, not real life. The term is still foggy but the main features chosen by respondents are: it happens online (63%), it is related to computers and the net (79%), it can be information distortion involving the government (65%), media (61%) and lay people (56%). Online piracy, massive unrest in the streets and rebellious activities on the net did not qualify.

In the next section we focused on the questions of who should be held responsible for cyber war and how the government should manage the problem. We found that governments are perceived as primary participants (90%). In addition, they are also blamed for individual people's acts (44%), especially when the attack comes from that location (47%). Governments should have the option to ask for help from international committees (86%), use other rights under the law (61%) or defend themselves using cyber defence tools (86%), EDLCDU (55%) or even with the help of individual hackers (41%). People's understanding and knowledge of related legislation is weak; 66% did not feel there is any legislation in that area at all. Also, more than 80% of the ICT and EDLCDU respondents claim to not have enough legal support in these issues.

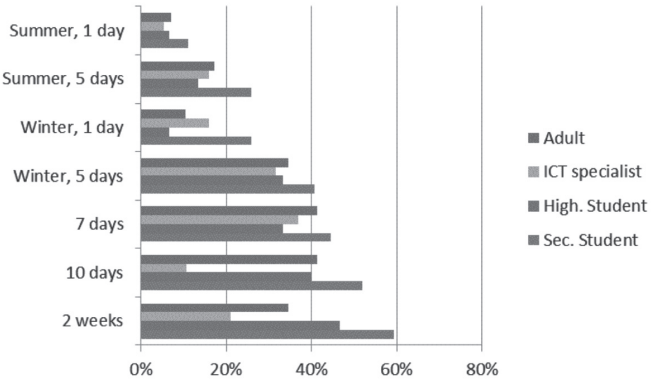
The most important thing that a government can do in a cyber war situation is to provide true, accurate and plentiful information. Massive unrest can be triggered when people find out that governments have manipulated information (61%) or restricted or filtered Internet usage (68%). People might understand if the government cuts some communication in that situation (75%) or gives informal groups special rights to regain control online (67%). The support for the EDLCDU and other supporting ICT cyber specialist groups is high (72%). The respondents would also like to have a government database of these specialists to ask for help from (77%).

For the third and fourth sub-categories we studied people's readiness to "go out to the streets" and possible chances to raise awareness in that area. Some interesting results were found among

the answers to the question "When internet, mobile network, electricity is unstable, ATM does not work, workplaces/ schools are closed, when would you start rallying in the streets or online (if possible)?" (see Figure 1). 100% of EDLCDU respondents answered this question that they would not act in any way that will cause more panic than there already is.

The difference between summer and winter was introduced to check people"s interest towards more peaceful solutions (go to the beach, visit grandparents or take a vacation). In summer there is no constant need for electricity and homes are warm, unlike in winter where it can reach -20 degrees etc. However, when something happens in the winter it is much more problematic. The difference can also be seen in the number of days of system downtime. The critical point is at days 5 to 7 when people would start going out onto the streets, before attacking shops to steal food. After the 10th day, some groups would start to find other solutions. After two weeks other adults' participation in rallying in the streets would also start to increase.

FIGURE 1. WHEN I WILL ACT?

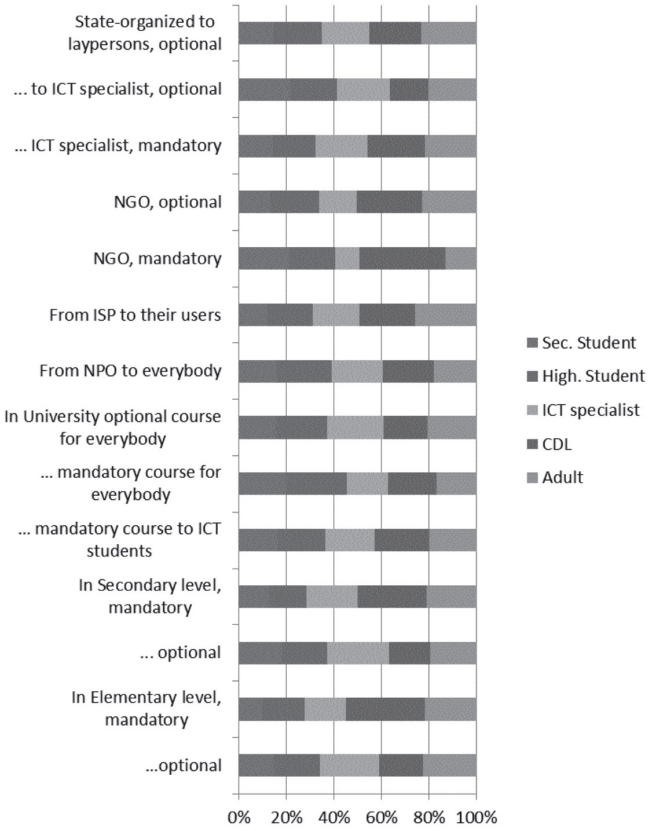


In the ethics section we focused on awareness training – where, when and how should it be carried out. 29% preferred theoretical training while the majority (71%) wanted a practical, hands-on approach. Practical training was relatively unimportant for secondary level students (59%). According to the respondents, the possibility of receiving practical training in cyber defence should be available to lay people (64%), CDL members (86%), university students (86%), secondary school students (68%) and elementary level students (40%). There were also implications that the training should not involve playing cyber war games (77%); the solution lies in building special information centres (54%) and events (53%) where people could be trained (e.g. in ICT security or programming). The perceived levels and locations are shown in Figure 2.

This kind of awareness is deemed to be the responsibility of the Estonian Ministries (and their affiliates): Economy and Communication (91%), Defence (86%) and Education (49%). However, the Estonian reality is that awareness in this area relies rather on the Ministry of Social Affairs (mentioned by 15%), the Tiger Leap Foundation (responsible for ICT and innovative methods implementation in Estonian schools; 10%), the children's welfare system (4%), and

volunteers (17%). It was also interesting that 24% of respondents expect their ISP and 17% the mass media to protect and advise them in this area.

FIGURE 2. CYBER WAR-RELATED EDUCATION – WHERE AND WHEN?



5. DISCUSSION

The results were somewhat expected – the ICT people have heard of cyber war, others see it as related only to "cyberspace and servers". At the same time, the most crucial finding seems to be that both the lack of information and (even perceived) manipulation attempts from the government will raise unrest in society.

We also see a problem when the media presents news originating from "an independent research agency" and is believed to be reliable a priori – only a few people will search more information about the topic or think about how newspapers translate original texts or "lend" coverage to each other (making it similar to the children's "telephone game"). It is also noteworthy that all information originating from Russia is considered by far the least reliable (compared to

the information from EU and the rest of the world). This can be at least partially explained by the fact that older generations learned to distrust everything Soviet (i.e. Russian) due to extensive (and often unreasonable) Communist propaganda during the occupation period, while the younger generations (especially the most active, Internet-savvy groups) developed almost the same level of distrust towards Russian information after the April 2007 riots.

Cyber war is in some ways like cyber bullying – it happens on the Internet and is often believed to be separate from real life. But as there are links between real-life bullying and cyber bullying, there are links between cyber war and real-life conflicts. When we see massive unrest in several countries, it has often started from discussions on the Internet where everybody can join – positive feedback takes place. It is easy to think that cyber incidents are related only to servers and systems, but it is important to see that these attacks are more and more related to information manipulation, lack of information and affecting people.

An interesting finding is the fact that 77% of the respondents would support the establishment of a central database of cyber defence specialists (the support was even seen among the EDLCDU and specialist groups). The idea is controversial at best, allowing any potential enemy to target a specific resource – and in case of success would compromise a large share of national cyber defence capability.

In Figure 1, we see children's dependence on online needs and quick reactions to problems. While adults tend to wait before acting and are more patient, youngsters usually want quick and simple results and are not patient. This can be illustrated by two events in Estonian history spanning over more than a dozen years – defending the TV Tower in 1991 [51] and the Bronze Soldier removal affair in 2007 – when schoolchildren and young adults were manifest in the streets. Young adults and students also formed the core of anti-ACTA protesters on February 11, 2012 [52].

It is interesting to see that by the 10th day, ICT specialists will start searching for other ways than demonstration. This may be due to the awareness of the different war games and scenarios, which ICT specialists have already faced; resource exploration, management etc. will be of prime importance.

The positive feedback that amplifies the oscillation may become problematic in such a situation, e.g. one person screams on the Internet and others will start to scream, then soon everybody will scream louder and louder. In that case rallying in the streets will happen faster and will involve more people.

When something related to spamming or phishing happens, people do not want to get involved, so usually they will not do anything, even if they see illegal things happening. When they see something unusual on TV (e.g. taking over the station) they will either dismiss it as a joke or search for additional information on the Internet to get confirmation.

We also see responsibility issues between awareness trainers – while one of the four Ministries is active, providing awareness training in the e-safety area, others are still wondering what to do or not to do. There is also a problem with projects which are imported from outside of Estonia

and do not fit into the society. More coordinated management is needed in this area.

By raising awareness of e-safety and cyber war issues, as well social engineering, less people will be affected by misleading information from any channels. Trusted and open government is also of vital importance. Awareness training should be carried out by specialists in that very area, not volunteers or different generic programmes. When people are treated respectfully by the government sharing valid information, there will be no or less rallying in the streets during the next crisis.

Education in cyber defence should be a part of national curricula at elementary, secondary and university levels. An interesting finding was that while people would like to have more practical skills to defend themselves, which is also seen as an opportunity to include NGO or ISP in the training process, government institutions were given preference over private enterprises. Establishment of information centres and the organisation of events like LAN-parties or training camps are also considered useful.

6. CONCLUSIONS

The respondents were willing to give up some rights during a crisis, but how it should be regulated perhaps needs an additional analysis. When the government needs help, they are happy to do that, but the request for assistance must be correct and well-addressed. The government must also consider that clear and abundant information during a crisis is very much appreciated.

The secondary level students and young adults were more willing to act in the streets and on the Internet when something negative occurs. Season and weather also play a role when considering the risk of public unrest. Due to the use of the Internet and social networking, simple events will get positive feedback (outburst) and they might create a snowball effect before calming down. These factors should be considered by SCADA specialists.

The responding experts also pointed out problems in the legislation and forensics area. On the one hand, there is a lack of awareness in that area; on the other hand there is a shortage of experts and funding to carry out these tasks, even when something serious happens. There is no official strategy and continuity yet to produce cyber security specialists.

The awareness training for common people should be a part of educational programmes in national curricula at secondary and university levels. It should be up to schools to decide how to execute it. The respondents also seemed to rely on government institutions to spread the "word", rather than trusting private enterprises to save the day. While the adults' training should be more practical, more theory is needed at the elementary level. Thus, we call for Estonian Ministries to further cooperate in raising cyber defence awareness among common people.

REFERENCES:

[1] K. Jõevere, (2011, Jan. 20) *Täna loodi Küberkaitseliit*, Eesti Päevaleht [onine] Avalilable: http://www.epl.ee/news/eesti/tana-loodi-kuberkaitseliit.d?id=51290517.

[2] A. Schaeffer-Filho et al *Future and Emerging Threats to Network Operation: A Quantitative Research Analysis*, Interim Report, Lancaster University, UK. Aug. 2011.

[3] B. Blunden, "Manufactured Consent and Cyberwar," in *LockDown Conference proceedings* University of Wisconsin-Madison, 2010.

[4] D. C. Rowe et al "The Role of Cyber-Security in Information Technology Education," in *ACM SIGITE 2011*, New York, USA, October 2011.

[5] B. West, "The Psychology of Security," *Communications of the ACM* Vol 51. No. 4 2008.

[6] R. Zheng et al "Effects of Motives for Internet Use, Aloneness, and Age Identity Gratifications on Online Social Behaviors and Social Support among Adolescents," in *Adolescent Online Social Communication and Behavior: Relationship Formation on the Internet*. Hershey, PA: IGI Global, Inc. 2008.

[7] T. Mataracioglu and S. Ozkan "User Awareness Mesaurment Through Social Engineering," in *Int. J. Managing Value and Supply Chains* 2010.

[8] M. Huber et al "Towards Automating Social Engineering Using Social Networking Sites," in *Computational Science and Engineering*, Vancouver BC. October 2009.

[9] J. R. C. Nurse et al "Trustworthy and Effective Communication of Cybersecurity Risks: A Review," *Socio-Technical Aspects in Security and Trust* (STAST), Nov. 2011.

[10] F. Stech et al "Scientometrics of Deception, Counter-deception, and Deception Detection in Cyber-space," in *PsychNology Journal* vol. 9 no.2 pp.79-122, 2011.

[11] R. Theobald, *The Rapids of Change: Social Entrepreneurship in Turbulent Times*. Knowledge Systems, Inc., Indianapolis, Indiana 1987, pp.82.

[12] R. Falkvinge, (2012, Feb. 4) *Nobody Asked for a Refrigerator Fee*. Falkvinge & Co on Infopolicy [online] Available: http://falkvinge.net/2012/02/04/nobody-asked-for-a-refrigerator-fee/.

[13] J. Naughton, "The internet: is it changing the way we think?" in *The Observer*, Aug. 2010 pp.20.

[14] T. R. Peltier, "Social Engineering: Concepts and Solutions," in *Auerbach Publications*, Nov. 2006.

[15] A. Raun, (2012, Feb. 7) *Lekkinud info: sajad našistid olid valmis Eestisse tulema*, Postimees. [online] Available: http://www.postimees.ee/731176/lekkinud-info-sajad-nasistid-olid-valmis-eestisse-tulema/.

[16] D. B. Farmer, *Do the Principles of War Apply to Cyber War?*, School of Advanced Military Studies United States Army Command and General Staff College Fort Leavenworth, Kansas 2010.

[17] G. V. Hulme, (2012 Jan. 4). *Government engineers actively plan for cyberwar*. CSOonline's Malware [online] Available: http://www.csoonline.com/article/697365/government-engineers-actively-plan-for-cyberwar.

[18] *Euroopa elutähtsate infrastruktuuride identifitseerimise ja määramise ning nende kaitse parandamise vajaduse hindamise kohta*, Nõukogu direktiiv 2011/114/EÜ, Euroopa Liidu Teataja pp 345/75.

[19] *Kapo: krooni devalveerimise paanika tekitajad teada* (2007, Dec. 13) Äripäev [online] Available: http://www.ap3.ee/?PublicationId=07d16439-8b41-4314-9fe6-e980bcaed68d.

[20] A. Lepa, "Eesti 2007. a. aprillirahutuste kajastamine Vene noorteorganisatsioon Na_i kodulehel" M.S. thesis, Dep. Phil. University of Tartu. Tartu, 2010.

[21] *Danish youths riot for 7th night, several arrested* (2008, Feb. 17) Reuters [online] Available: http://in.reuters.com/article/2008/02/17/idINIndia-31995320080217.

[22] *London riots: Looting and violence continues* (2011, Aug. 8) BBC News London [online] Available: http://www.bbc.co.uk/news/uk-england-london-14439970.

[23] J. Lichfield, (2010 Oct. 19). *France braces for riots as protests turn violent*, The Independent [online] Available: http://www.independent.co.uk/news/world/europe/france-braces-for-riots-as-protests-turn-violent-2110305.html.

[24] L. Wood, The Arabian Spring and its Impact on MENA Economies, *Research and Markets*, Dec 2011.

[25] M. Lennon, (2011, Mar. 11). *Massive Influx of Scams Surrounding Japan's Earthquake and Tsunami Expected*. Security Week [online] Available: http://www.securityweek.com/massive-influx-scams-surrounding-japans-earthquake-and-tsunami-expected.

[26] J. Ugander, et al "The Anatomy of the Facebook Social Graph," *Cornell University Library*, NY, US. 2011.

[27] L. Backstrom, et al, "Four Degrees of Separation," *Cornell University Library*, NY, US. 2011.

[28] T. Goetz, "Harnessing the Power of Feedback Loops," *Wired Magazine*, Jun. 2011.

[29] P. McRae, "Argumentum ad infinitum: The complex nature of echoing voices on the Internet," *Complexity Science and Educational Research (CSER)*, Vancouver, BC: University of British Columbia. 2007.

[30] K. Lewis, (2010, Feb 11) *The Network Effect* [online] Available: http://www.kieranlevis.com/the-network-effect/.

[31] L. Rosales, (2011, Oct 18) *Anatomy of a social media chain reaction – case study* [online] Available: http://agbeat.com/real-estate-technology-new-media/anatomy-of-a-social-media-chain-reaction-case-study/.
[32] D. Paap, (2009, Aug. 3) *Dance Tributes Around the World for the Dance Legend, Michael Jackson* [online] Available: http://movetheframe.wordpress.com/2009/08/03/dance-tributes-around-the-world-for-the-dance-legend-michael-jackson/.
[33] *Hundreds of Occupy Wall Street protesters arrested* (2011, Oct. 2) BBC News US&Canada [online] Available: http://www.bbc.co.uk/news/world-us-canada-15140671.
[34] *Acta: Europe braced for protests over anti-piracy treaty* (2012 Feb. 6) BBC News Technology [online] Available: http://www.bbc.co.uk/news/technology-16906086.
[35] E. G. Cohn, "Weather ans Crime," *Brit. J. Criminol* vol 30 no. 1. 1990.
[36] P. Butke and S. Sheridan, "An Analysis of Relationships between Weather ans Aggressive Crime in Cleavland, Ohio," Dep. Geography, Ken State Univerity, 2010.
[37] K. Leetaru, "Cultoromics 2.0: Forecasting large-scale human behaviour using news media tone in time and space," *FM* Vol 16 no 9 2011.
[38] B. Schaller and R. Stough, "The Origin, Nature, and Implications of MOORE'S LAW," *PUBP801*, Sept. 1996.
[39] K. Jaroslaw, "Civiliziting events ans chronology," *Proceedings of the 2nd International Meeting A Revised Chronology and Alternative History*, Rüspe, Germany, June, 2001.
[40] *Politsei süüteoennetusliku tegevuse 2011. aasta plaan* [online] Available: http://www.politsei.ee/dotAsset/174500.pdf 2011.
[41] *Kuritegevus Eestis 2010*, Kriminaalpoliitika uuringud 15 [online] Available: http://www.just.ee/orb.aw/class=file/action=preview/id=54601/KuritegevusEestis2010_web.pdf 2010.
[42] *Kriminaalpoliitika arengusuunad aastani 2018* [online] Available: https://www.riigiteataja.ee/akt/13329831 2010.
[43] J. Boyd, "Information Warfare OODA loop," in *Value Based Management* 2003.
[44] K. Saalbach, "Cyber war Methods ans Practice," *LV Internet policy Universität Osnabrück* Jan 2011.
[45] A. Veldre, "E-ühiskond," *Pühajärve suvekool* [online] Available: xyz.ee/2011-08-23-suvekool Aug. 2011.
[46] E. Tõugu, (2008 Jun. 17) *Kompetentsikeskus on, kus on kompetents?* Eesti Päevaleht [online] Available: http://www.epl.ee/news/melu/kompetentsikeskus-on-kus-on-kompetents.d?id=51133361.
[47] *About the project Targalt Internetis* (2010, Oct. 28) [online] Available: http://www.targaltinternetis.ee/projektist/?lang=en 2010.
[48] *Ole Kaasas! Eesmärk* (2011, Feb. 12) [online] Available: http://www.olekaasas.ee/eesmark/ 2011
[49] R. Kumar, *Research Methodology*, APH Publishing, 2005.
[50] Z. Todd, *Mixing methods in psychology: the integration of qualitative and practice*, Psychology Press Taylor ans Francis Group, 2004.
[51] *Saatuslikud tunnid teletornis?* (2006 Aug. 17) Maaleht [online] Available: http://www.levira.ee/dyna/site/696est.html.
[52] *Tartu ACTA vastaste protesti lõpukõne: Siim Tuisk* (2012, Feb. 11) Youtube [onlne] Available: http://www.youtube.com/watch?v=gDz8MAG4jj4.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

The Notion of Combatancy in Cyber Warfare*

Sean Watts
Creighton University Law School
Omaha, Nebraska, U.S.A.
United States Military Academy at West Point
West Point, U.S.A.

Abstract: The class of combatant constitutes one of the most important instrumentalities of the law of war. Combatant status resolves critical and enduring legal questions such as immunity from prosecution for warlike acts, susceptibility to intentional targeting, and, in part, treatment upon capture. Since the late nineteenth century, codifications of the international law of war have included criteria for combatant status keyed to ensuring desirable battlefield conduct and, to the extent possible, humanity in war. This paper revisits the author's prior work on the topic of combatancy in cyber warfare. Building on recent public revelations concerning state capacity for offensive cyber attacks, as well as new developments in computer network attack, this paper highlights logical and normative shortcomings in current understandings of combatant status in cyberspace. In place of rote reliance on existing criteria intended for the kinetic battlefield, this paper proposes reliance on State affiliation as the sole criterion for evaluating combatant status in cyber warfare between States. An admitted interpretive gloss on current criteria, the proposed framework offers a workable and realistic reconciliation of humanitarian goals and emerging State practice in cyber warfare.

Keywords: *International Humanitarian Law, Law of Armed Conflict, Law of War, cyber attack, cyber warfare, combatant status*

1. INTRODUCTION

The laws of war occasionally paint an idealized portrait of armed conflict. An impression of war formed exclusively from the international legal instruments that regulate the conduct of hostilities would render an image perhaps foreign to present day combatants. In lieu of surprise attacks, one would find punctilious declarations of hostilities in the form of diplomatic notes or ultimatums.[1] States would investigate detention conditions and communicate their humanitarian

* This paper updates concepts and ideas developed previously and in greater depth in *Combatant Status and Computer Network Attack*, 50 Virginia Journal of International Law 392 (2009)[hereinafter Watts].

1 Convention Relative to the Opening of Hostilities, art. 1, Oct. 18, 1907, 36 Stat. 2259, 1 Bevans 619 [hereinafter 1907 Hague Convention III](requiring that contracting Powers not commence hostilities "without previous and explicit warning).

concerns to one another through mutually acceptable third State parties.[2] Combatants would make themselves physically separate and visually distinguishable from civilians through the wear and display of distinctive uniforms and insignia.[3] Civilian populations would be warned in advance of impending attacks and provided an opportunity to evacuate to safe areas.[4] Safety zones, immune from attack would be created in the midst of the battlefield to provide shelter to children and the elderly.[5] And belligerents would facilitate the transport of wounded by air through agreed flight plans for medical aircraft, even through enemy territory.[6]

The reality of modern warfare is, of course, quite different. States rarely resort to declarations of war any longer. The Geneva Conventions' Protecting Power scheme almost never operates through third party States. The modern battlefield sees fighters intermingled with and often indistinct from their civilian counterparts. The opportunity to warn civilians of impending bombardments or attacks, without dooming such operations to failure, rarely presents itself. Civilians are all-too-often caught up in or the object of military attacks. And, as yet, agreements between belligerents permitting enemy medical aircraft to fly over friendly-controlled territory have not become standard operating procedure.

Yet it is too much to say that the law of war is entirely irrelevant or ineffectual. It is still probably correct to say that most States regard the law of war as more than merely epiphenomenal. In fact, States' militaries and government agencies have largely internalized and rendered operational the great majority of the present laws of war. For example, the proliferation of serious military legal manuals provides doctrinal evidence that States regard the law of war as relevant and meaningfully binding.[7] In an era when many armed forces face personnel cuts, reliance on sizable corps of military and civilian lawyers to review and advise on planning and operations reflects States' real commitment to the notion of legal restraint in war. And prosecutions at international criminal tribunals reflect both States' willingness to dedicate significant resources to the law as well as their commitment to enforce at least the principles, if not always the exact

2 Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field art. 8, Aug. 12, 1949, 6 U.S.T. 3114, 75 U.N.T.S. 31 [hereinafter 1949 Geneva Convention I]; Convention for the Amelioration of the Condition of the Wounded, Sick, and Shipwrecked Members of Armed Forces at Sea art. 8, Aug. 12, 1949, 6 U.S.T. 3217, 75 U.N.T.S. 85 [hereinafter 1949 Geneva Convention II]; Convention Relative to the Treatment of Prisoners of War art.8, Aug. 12, 1949, 6 U.S.T. 3316, 75 U.N.T.S. 135 [hereinafter 1949 Geneva Convention III]; Convention Relative to the Protection of Civilian Persons in Time of War art. 9, Aug. 12, 1949, 6 U.S.T. 3516, 75 U.N.T.S. 287 [hereinafter 1949 Geneva Convention IV](concerning the appointment of Protecting Powers by parties to an international armed conflict for purposes of implementing the 1949 Geneva Conventions).

3 1949 Geneva Convention III, *supra* note 2, art 4A(2)(b); Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, art. 44(7), June 8, 1977, 1125 U.N.T.S. 3 [hereinafter Protocol I](incentivizing display distinctive insignia and wear of uniforms by combatants through conferral of prisoner-of-war status).

4 Protocol I, *supra* note 3, art. 57(2)(c); Convention Respecting the Laws and Customs of War on Land, Annex to the Conventions, art. 26 Oct. 18, 1907, 36 Stat. 2277, 1 Bevans 631 [hereinafter 1907 Hague IV Regulations].

5 1949 Geneva Convention IV, *supra* note 2, art. 14.

6 Protocol I, *supra* note 3, arts 26 & 29.

7 *See e.g.* United States Department of the Navy, The Commander's Handbook on the Law of Naval Operations, NWP1-14M, (July 2007); United Kingdom Ministry of Defence, The Manual of the Law of Armed Conflict (2004); Canadian Office of the Judge Advocate General, Law of Armed Conflict at the Operational and Tactical Levels (Aug. 13, 2001); Federal Republic of Germany, Federal Ministry of Defence, (Aug. 1992). A study of customary international laws of war by the International Committee of the Red Cross draws on a far broader sampling of States' law-of-war manuals. 2 Jean-Marie Henckaerts & Louise Doswald-Beck, Customary International Humanitarian Law (2005).

letter, of existing international law regulating the conduct of hostilities.[8]

Owing to disparity between the letter of the law of war and its practical implementation, much of States' adherence to the law of war relies on secondary rules.[9] So often drafted long before the evolutions and revolutions in the wars they regulate, law-of-war treaties frequently require adaptive understandings, interpretive canons, or operational implementation – so-called secondary rules – to remain relevant. In addition to supporting claims of general legal efficacy, military manuals, military lawyers, and international criminal trials each contribute to States' efforts to adapt existing law to the evolving realities of armed conflict. Law of war manuals contextualize obligations, providing interpretation and examples of implementation. Military lawyers operationalize legal principles and rules through advice during planning and execution of orders, adapting law to battlefield conditions and evolving threats. And tribunals, domestic and international, perform a similar function, interpreting and applying law-of-war terms, often according to their perceived object and purpose. In short, although faced with idealized expressions and often dated assumptions, States continue to honor the law of war aided by a variety of interpretive measures. Interpretation vindicates the law by ensuring its relevance and vitality, operationalizing humanitarian ideals to the extent possible while assuring military effectiveness and realism.

This paper seeks briefly to illustrate and defend such an interpretation in the context of an emerging and revolutionary form of warfare – cyber war. This paper will briefly address the important question of combatancy or combatant status in cyber warfare. In particular, the question of who may directly participate in cyber hostilities will be addressed. If States have developed a class of cyber warriors, must they be drawn from or incorporated into regular armed forces? Or may a State sanction and employ civilian actors to conduct cyber attacks and other warlike operations in cyberspace?

Like many law-of-war provisions, the criteria for combatant status are derived from long-standing traditions. Chosen both to reflect and to reinforce classic attributes of legitimate belligerents, the combatant criteria perform gate-keeping functions for both prisoner of war status, and immunity from prosecution for lawful warlike acts as well as the critical question of exposure to intentional targeting. While well-suited to the battlefields of centuries past, I argue that the traditional combatant criteria are applied over-broadly to participants in emerging forms of remote warfare such as computer network warfare. Increasingly these rules misapprehend how and, more importantly, by whom modern war such as cyber warfare will likely be fought.

This paper proposes an alternate test for combatant status in cyber warfare focused on State affiliation. Long an important, yet overlooked criterion for combatant status, State affiliation enjoys solid textual support in the extant law and supports the fundamental principles of distinction and discipline through State responsibility. But perhaps most importantly, State affiliation as a criterion for lawful combatancy in cyber warfare is minimally disruptive to emerging State practice thus guaranteeing relevance and alignment of the law with the realities

8 *See* Prosecutor v. Gotovina et al., IT-06-90-T, 15 Apr. 2011 (sentencing two senior Croatian military officers to 24 years and 18 years confinement for indiscriminate artillery shelling and a joint criminal enterprise to persecute and deport ethnic Serbians).

9 The English legal philosopher H.L.A. Hart identified secondary rules as rules that give effect to primary rules that directly regulate conduct. Rules of adjudication, interpretation, and that prescribe the operation of primary rules constitute secondary rules. H.L.A. Hart, The Concept of Law, 77-79, 88-93 (1961).

of the cyber battlefield. State affiliation as a stand-alone sole criterion is admittedly a gloss on the present law of combatant status, perhaps at this point more in the nature of *lex ferenda*. However it is an interpretation that overcomes the existing law's static and dated character, augmenting its legitimacy by reconciling what States say with what States actually do and will do in cyber warfare.[10]

2. THE INTERNATIONAL LAW OF COMBATANCY

In contrast to its public international law cohort, international human rights law, the law of war has long relied on classifications to allocate protections, duties, and responsibilities.[11] Where the protections of human rights law apply merely by virtue of personhood, law-of-war protections have generally been contingent upon persons' satisfaction of particular criteria, such as nationality, membership in an organization, or a prescribed course of conduct. Presently, the most important law-of-war classifications with respect to persons are the civilian and combatant classes. This section briefly outlines the traditions, legal framework, and consequences of the law-of-war status of combatant.

The earliest attempts to draft multilateral law-of-war treaties recognized the status of combatant, beginning with the 1874 Brussels Declaration.[12] Designed to capture the customs and usages of militaries that alleviated unnecessary suffering in war, the Declaration applied the "laws, rights, and duties of war . . . not only to armies, but also to militia and volunteer corps fulfilling the following conditions:

1. That they be commanded by a person responsible for his subordinates;
2. That they have a fixed distinctive emblem recognizable at a distance;
3. That they carry arms openly;
4. That they conduct their operations in accordance with the laws and customs of war."[13]

The Declaration's description of the combatant class was noteworthy in several respects. First, the Declaration's definition was an expansive conception of the combatant class. The definition included not merely States' regular armed forces but also irregular or mustered volunteers. It was at once progressive and conventional. The definition would give international recognition and legal status to emergent fighting forces, yet by qualifying their combatant status on satisfaction of the four enumerated criteria, the Declaration incentivized conformity with the traditional behaviors, appearances, and customs of States' regular armed forces. Since the Declaration was drafted, States have continued to debate the merits of legal recognition of unconventional fighting organizations. Yet as recently as 2002, States have identified the four criteria as essential attributes of organized armed forces, including a controversial U.S. legal opinion requiring that even regular armed forces fulfill the four 1874 criteria to legitimately

10 *Id*.

11 Although dispute exists as to the geographic applicability of many human rights norms and treaties, once activated human rights obligations are generally accepted as universally applicable to all persons, regardless of citizenship, national origin, or political alliance.

12 Project of an International Declaration Concerning the Laws and Customs of War, Aug. 27, 1874, 4 Martens Nouveau Recueil (ser. 2) 219 [hereinafter 1874 Brussels Declaration]. Despite its seemingly fundamental protections, the Declaration appears to have gone too far for most of its signatories as it never entered force. *See* The Laws of Armed Conflicts 21 (Dietrich Schindler & Jiri Toman, eds., 2004).

13 1874 Brussels Declaration, *supra* note 9, art. 9.

claim combatant status.[14]

The Declaration's description of the combatant class is also noteworthy for its longevity. Although the Declaration never entered into force itself, succeeding multilateral law-of-war treaties liberally incorporated its definition. The 1899 Hague Convention II,[15] the 1907 Hague Convention IV Annexed Regulations,[16] the 1929 Geneva Prisoners of War Convention,[17] and the 1949 Third Geneva Convention[18] all reproduce or incorporate the 1874 criteria by reference in their descriptions of combatants. With the important exception of a clearer reference to the requirement of State affiliation in the 1949 Third Geneva Convention, the 1874 criteria operated nearly unchanged for over 100 years.[19] Not until 1977, with Additional Protocol I to the Geneva Conventions, did the international law of war tinker with the 1874 Declaration's formula for combatant status. Yet even Additional Protocol I remained grounded in the 1874 criteria to a significant extent.

Polemical accounts criticize Additional Protocol I for rendering meaningless the class of combatant.[20] Such critiques focus on the Protocol's abandonment of the traditional combatant criteria. It is true that the Protocol's modification of the 1874 criteria drew significant dissent, including a number of reservations by States Parties,[21] as well as refusals to ratify by States attending the diplomatic conference.[22] Closer examination, however, reveals the persistent, though marginally reduced, influence of the 1874 criteria.

Additional Protocol I defines combatants as "[m]embers of the armed forces of a Party to a conflict [...]"[23] Elaborating on the term "armed forces" the Protocol adds,

> "The armed forces of a Party to a conflict consist of all organized armed forces, groups and units which are under a command responsible to that Party for the conduct of its subordinates, even if that Party is represented by a government or an authority not recognized by an adverse Party. Such armed forces shall be subject to an internal disciplinary system which, inter alia, shall enforce compliance with the rules of international law applicable in armed conflict."[24]

14 Memorandum from Jay S. Bybee, Assistant Att'y Gen., Office of Legal Counsel, Dep't of Justice, to Alberto Gonzales, Counsel to the President, and William J. Haynes II, Gen. Counsel of the DOD, *Application of Treaties and Laws to al Qaeda and Taliban Detainees* 10 (Jan. 22, 2002) in The Torture Papers (Karen J. Greenberg & Joshua L. Dratel eds., 2005).

15 Convention with Respect to the Laws and Customs of War on Land, art. 1, July 29, 1899, 32 Stat. 1803, 26 Martens Nouveau Recueil (ser. 2) 949.

16 1907 Hague IV Regulations, *supra* note 4, art. 1.

17 Convention Relative to the Treatment of Prisoners of War, art. 1(1), July 27, 1929, 47 Stat. 2021, 118 L.N.T.S. 343.

18 1949 Geneva Convention III, *supra* note 2, art. 4A(2).

19 *Id*. (prefacing the four 1874 criteria with, "Members of militias and members of other volunteer corps, including those of organized resistance movements *belonging to a Party to the conflict* [...].")(emphasis added).

20 Douglas J. Feith, *Law in the Service of Terror*, The National Interest (Fall 1985).

21 *See* United Kingdom Reservations to Additional Protocol I to the Geneva Conventions (July 2, 2002), *available at* http://www.icrc.org/ihl.nsf/NORM/0A9E03F0F2EE757CC1256402003FB6D2?OpenDocument.

22 *See e.g*. Letter of Transmittal and Letter of Submittal Relating to Protocol II Additional to the Geneva Conventions of 12 August 1949 (Jan. 29, 1987), *reprinted in* U.S. Dep't of the Navy, Annotated Supplement to the Commander's Handbook on the Law of Naval Operations, 306 (A.R. Thomas & James C. Duncan eds., 1999).

23 1977 Additional Protocol I, *supra* note 3, art. 43(2).

24 *Id*. art. 43(1).

The influence of the 1874 criteria is obvious. Even under the Additional Protocol's relaxed rules for combatancy fighting organizations must still appoint and take direction from a superior commander and must conform their conduct of hostilities generally to the law of war. In fact, Additional Protocol I only departs from two of the four 1874 criteria and does so only in a limited sense. In fact, article 44 requires that combatants "distinguish themselves from the civilian population" and "carr[y] arms openly." Facially, the latter requirement with respect to carrying arms makes no change to the traditional rule. The former requirement with respect to distinction, although abandoning the 1874 phraseology, also performs substantially the same function as its forebear. Rather than dispense with uniforms and military insignia entirely, the Protocol's phrasing merely seems to admit alternate visual indicia of fighting organizations' hostile function, such as clothing or armbands.[25]

The only notable Additional Protocol I alteration to the 1874 criteria concerns a limited exception for guerilla fighters and insurgent groups in enemy-occupied territory. Article 44 relaxes the distinction and arms criteria when "owing to the nature [...] [of] hostilities," observance would be impracticable.[26] The exception is not available during attacks or when visible to enemy forces while preparing for or deploying to attack. Concerned with the negative implications for civilian populations, the majority of delegations to the Additional Protocol's diplomatic conference understood the exception to be limited to non-combat related movements in occupied territory.[27] In the vast majority of circumstances related to combat, the four 1874 criteria operate under Additional Protocol I has they had for over a century. Thus, in the majority of circumstances even Additional Protocol I preserves the four 1874 criteria as the essential prerequisites to combatant status.

The 1874 Declaration and its criteria are also remarkable for their attention to the realities and demands of late nineteenth and early twentieth century warfare. Each criterion performed an important function in ensuring warfare between States was distinguishable from uncontrolled violence. The first criterion, the responsible command requirement, ensured that lawful participation in warfare was limited to organized groups operating on behalf of States. The responsible command function excluded individual opportunists, criminals, and brigands from combatant status. Additionally, the command criterion aided accountability and adherence to law, ensuring superiors presumably better steeped in the traditions and customs of lawful combat supervised their combatants' actions. One found on the battlefield, and one often still finds today, an environment ripe for criminal exploitation. Suspended civil capacity, vulnerable and displaced populations, damaged or abandoned property, and general chaos present convenient conditions for looting, rape, and other criminal activity. Additionally, in war, individual armed belligerents often wield power out of proportion to their authority. Command and the attendant systems of internal discipline emblematic of armed forces stood as essential deterrents to battlefield bedlam. Military command structures, with their strict hierarchies and rigorous lines of authority, operated effectively despite physical and geographic separation between the leader and led. Military command ensured that combatants limited their conduct to actions that were militarily necessary.

The second and third of the 1874 criteria, that combatants wear distinctive emblems and carry

25 *See* Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949, 527-28 (Yves Sandoz et al. eds., 1987)[hereinafter Additional Protocol Commentary].
26 *Id*. art. 44(3).
27 *See* Additional Protocol Commentary, *supra* note 20, at 530-32.

arms openly, also performed an important battlefield function related to humanity and civility. Uniforms and the open display of military arms greatly facilitated opposing forces' efforts to distinguish enemy combatants from civilians. With engagements limited to visual range, displays of distinctly military uniforms and weapons were a particularly effective means of limiting the effects of hostilities to combatants on the late nineteenth century battlefield. Nor did the twentieth century's widespread use of beyond-visual-range or over-the-horizon weapons render the uniform and arms criteria useless. Line-of-sight engagements remained prevalent features of twentieth century kinetic armed conflict. Moreover, forward observers or other combatants directing and adjusting the fire of indirect and over-the-horizon weapon systems could still rely on uniforms and the open display of weapons to distinguish lawful targets from protected civilians.

Last, the requirement that combatants' organizations conform their conduct to the laws and customs of war performed an important reinforcing function. A form of reciprocity, the fourth 1874 criterion excluded from the combatant class groups of fighters unwilling to adhere to traditional and recognized limits on the conduct of hostilities. Members of groups regularly resorting to perfidy, treachery, indiscriminate attack, use of prohibited weapons, or maltreatment of victims of war could not claim the law's protections accorded to combatants upon capture. Requiring that combatant organizations conduct their operations in accordance with the law of war also incentivized individual instruction in the law to guarantee continued combatant status and its attendant protections and privileges. Physically separated from and often out of communication with legal advisors and senior leaders, nineteenth and twentieth century combatants could be distinguished from their unlawful belligerent counterparts for the internal familiarity with and general observance of the rudiments of lawful battlefield conduct.

No explanation of combatancy under the law of war would be complete without discussion of its functions. Like all forms of status under the law of war, combatant status is a legal instrumentality – a means of prescribing and allocating legal obligations and protections. In short, three consequences flow from assignment of combatant status – only one of which is exclusive to that class.

The most important and the only exclusive consequence of combatant status is immunity from prosecution for lawful warlike acts. It is widely accepted that combatants may not be brought to criminal trial for acts of destruction or killings they commit in war.[28] Although combatant immunity (also known as the combatant's privilege) is well-established in the customs of war, the principle appeared relatively late in the codified laws of war. Additional Protocol I of 1977 appears to be the first multilateral codification of combatant immunity, providing, "Members of the armed forces [...] have a right to participate directly in hostilities."[29] While debate exists whether direct participation in hostilities by persons not qualifying as combatants constitutes an individual criminal offense under international law, it is quite clear that neither international nor domestic criminal tribunals may prosecute the otherwise lawful warlike acts

28 Anicee van Engeland, Civilian or Combatant? A Challenge for the 21st Century, 45 (2011); Knut Ipsen, *Combatants and Non-Combatants, in* The Handbook of Humanitarian Law in Armed Conflicts 81 (Dieter Fleck ed., 1995).

29 1977 Additional Protocol I, *supra* note 3, art. 43(2).

of combatants.[30] Acts of combatants that violate discreet law-of-war rules are punishable, such as perfidy, indiscriminate attack, use of unlawful weapons or means of war, or maltreatment of protected persons. However, the mere fact of combatants' direct participation in hostilities itself is privileged and perhaps the most significant by-product of combatant status.

A second consequence of combatant status is conferral of prisoner of war status upon capture. The concept of prisoner of war is ancient and has included progressively comprehensive protections as the law-of-war has developed.[31] In general, captors may only impose restraints on the liberty of prisoners of war necessary to prevent their return to the battlefield. Properly carried out, prisoner of war detention has more in common with camp or internment settings than with criminal incarceration. Prisoners of war are guaranteed payment, protection from abuse, recreational opportunities, limits on forced labor, significant procedural protections from discipline and punishment, communication with family members, and regular medical treatment.[32] Upon termination of hostilities, detaining powers must repatriate prisoners of war to their countries of origin. Unlike combatant immunity, prisoner of war status is not exclusive to combatants. At least two classes of civilians are also entitled to prisoner of war status upon capture: contractors, correspondents, and laborers accompanying the armed forces; and crews of merchant marine ships and civil aircraft used by belligerents.[33]

The final significant consequence of combatant status is exposure to status-based targeting by enemy forces. Combatants are lawful targets for their enemies' operations at all times until their surrender, capture, or incapacitation by wounds.[34] It is their status as combatants, their formal affiliation with and conduct of hostilities on behalf of an enemy State in international armed conflict, rather than their conduct that makes combatants lawful targets. Whether a combatant is in uniform or not, on duty or not, conducting an attack, or sleeping, she is a lawful target for enemy forces. Classically, status-based susceptibility to targeting has been a condition unique to the combatant class.[35] While civilians are subject to lawful targeting while taking direct part in hostilities, they are only lawful targets "for such time as" or while they actually commit hostile acts directly producing harmful effects to an enemy.[36] In this respect hostile civilians can be said to be targetable only on the basis of their conduct rather than any status. However, recently the exclusivity of combatants' status-based exposure to targeting been challenged widely.[37]

Thus, combatant status constitutes a central and remarkably static feature of the regulation of hostilities. From the time when war featured massed formations of distinctly-clad soldiers

30 *See* Richard R. Baxter, *So-Called 'Unprivileged Belligerency': Spies, Guerillas, and Saboteurs*, 28 Brit. Y.B. Int'l L. 323 (1951); Knut Dörmann, *The Legal Situation of "Unlawful/Unprivileged Combatants*," 85 Int'l Rev. Red Cross 45 (2003). For discussion of whether direct participation in hostilities by persons not qualifying for combatant status constitutes a crime under the international law of war see Mark David 'Max' Maxwell & Sean Watts, *'Unlawful Enemy Combatant': Legal Status, Theory of Culpability, or Neither*, 5 Journal of International Criminal Justice 19 (2007).

31 For an exceptionally thorough treatment of prisoner of war status, see 59 International Law Studies: Prisoners of War in International Conflict (Howard S. Levie, ed., 1979).

32 See 1949 Geneva Convention III, *supra* note 2, Part III.

33 1949 Geneva Convention III, *supra* note 2, art. 4(A)(4) & (5).

34 *See* Yoram Dinstein, The Conduct of Hostilities under the Law of International Armed Conflict, 34 (2d ed., 2010).

35 A recent study sponsored by the International Committee of the Red Cross with growing international support appears to extend status-based targeting to members of so-called organized armed groups. Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law, 31-35 (2009).

36 1977 Additional Protocol I, *supra* note 3, art. 51(3).

37 See Interpretive Guidance, *supra* note 35.

facing off with short-range rifles to the age of transcontinental missiles and remotely-piloted attack drones, relatively little with respect to the legal qualifications for or consequences of combatancy has changed. The following section inquires whether the static nature of combatancy is appropriate in light what is known and expected to develop in the emerging forms of conflict such as cyber warfare.

3. COMBATANCY IN CYBER WARFARE

In a prior article addressing the topic of combatant status and computer network attack, I used incidents in Estonia in 2007 and in Georgia in 2008 to illustrate the nature and effects of hostile computer network operations.[38] For authors addressing the legal aspects of cyber warfare at that time, the Estonian and Georgian directed denial of service incidents offered the most prominent, publicly available examples of international computer network incidents intended to harm States. Yet each incident offered minimal assistance in illustrating the operation of law-of-war principles in the cyber context. As most experts would agree, neither incident on its own constituted an "attack" for purposes of the law of war. Viewed alone, each likely amounted to a mere disruption of communications or inconvenience. At best, the Estonian and Georgian incidents illustrated the likelihood that States could impose significant disruptions through cyber means and would likely dedicate significant resources in the future to developing and countering cyber capacity to carry out cyber operations that might truly amount to attacks in the legal sense.

Since the Estonian and Georgian incidents, two developments have better framed the realities of computer network attack (CNA). First, one need no longer speculate or read between the lines of budget requests, as I did earlier, to determine whether States possess offensive cyber capacity. States have made clear that cyberspace is an important military domain.[39] Some States have even publicly acknowledged their capacity for offensive cyber operations amounting to attack.[40] A recent United States Defense Authorization Act curiously includes the following, "Congress confirms that the Department of Defense has the capability, and upon direction by the President may conduct offensive operations in cyberspace [...]."[41] And in a 2011, statutorily required report to the United States Congress, the Department of Defense revealed publicly, "[T]he Department has the capability to conduct offensive operations in cyberspace to defend our Nation, Allies and interests. If directed by the President, DoD will conduct offensive cyber operations in a manner consistent with the policy principles and legal regimes that the Department follows for kinetic capabilities, including the law of armed conflict."[42] Thus, it is

38 See Watts, *supra* note *, at 397-407.

39 *See e.g.* United States Department of Defense, *Strategy for Operating in Cyberspace*, 5 (2011) (resolving to treat cyberspace as operational domain).

40 *See* Uzi Mahnaimi, *Israeli Military Plots to Cripple Iran in Cyberspace*, London Sunday Times (Aug. 7, 2011) (describing an Israeli military cyber command reporting directly to the Prime Minister)[herinafter Mahnaimi].

41 2012 National Defense Authorization Act, sec. 954.

42 United States Department of Defense, *Cyberspace Policy Report: A Report to Congress Pursuant to the National Defense Authorization Act for Fiscal Year 2011*, Section 934, 5 (Nov. 2011) [hereinafter *Cyberspace Policy Report*].

clear that States have developed internal capacity, including cadres of cyber warriors, dedicated to conducting network warfare.[43]

A second development confirming the nature and extent of hostile cyber capacity is the 2010 Stuxnet worm attack on Iran. Discovered publicly in July of 2010, Stuxnet was a complex, malicious code believed to have been designed and introduced to sabotage industrial control systems in the Iranian nuclear program.[44] Combining an array of at least nine distinct variants of malware, including four invaluable zero day exploits, Stuxnet first infected Windows-based computers then spread to others in search of its target industrial control systems.[45] Although initially introduced to relatively few systems, Stuxnet later self-replicated to affect many more target systems.[46] It also appears the creators of Stuxnet updated and improved the worm as the attack unfolded. Earlier infected systems even requested and received updated versions of Stuxnet.[47] Once embedded in its final target system, Stuxnet modified and provided faulty performance feedback to control systems causing those systems to issue destructive operating commands to the machines they controlled.[48] It is estimated that Stuxnet caused sufficient physical damage to Iranian nuclear industrial apparatuses to set the program back one to two years.[49] To many, the unprecedented sophistication of the operation suggested that only State actors could have launched the attack.[50]

More so than previously revealed cyber operations, Stuxnet illustrates the potential of cyber operations to rise the level of attack under the law of war. If the hallmark of attack under the law of war is physically destructive effects, Stuxnet clearly qualifies. Stuxnet makes clear that crippling and physically destructive attacks on critical infrastructure are entirely possible and not merely the imaginings of worst-case scenario doomsayers. From events such as the Stuxnet attack it is also clear that destructive CNAs are complex, multi-stage operations. Analysts have concluded that the Stuxnet attack featured many of the attributes of conventional military operations including intelligence operations and mid-operation fragmentary orders. The attack involved a significant reconnaissance effort, likely including earlier intrusions into target systems.[51] Intelligence details that would have been useful to CNA operations such as Stuxnet include physical configuration of hardware, Internet Protocol addresses of connected computers, security patch installation histories, target platform operating systems, operator identities, and information on delivery of computer components to the target facility.[52] As noted above, rather than simply operating as off-the-shelf code, Stuxnet appears to have been designed, updated and even manipulated by it operators during the attack. It also appears the operation was monitored and commanded while in progress as are conventional, kinetic military operations.

43 In May 2010, the United States Department of Defense activated the U.S. Cyber Command, a military organization devoted to cyber operations. See Ellen Nakashima, *Gates Creates Cyber0Defense Command*, Washington Post, Jun. 24, 2009, at http://www.washingtonpost.com/wpdyn/content/article/2009/06/23/AR2009062303492.html.

44 David E. Sanger, *Iran Fights Malware Attacking Computers*, New York Times (Sep. 26, 2010).

45 Nicolas Falliere, et al., *W32.Stuxnet Dossier, Version 1.3*, Symantec Security Response, 1-2 (Nov. 2010) [hereinafter Falliere et al.].

46 *Id*. at 21.

47 *Id*.

48 *Id*.

49 David E. Sanger, *America's Dearly Dynamics With Iran*, New York Times (Nov. 6, 2011).

50 Mahnaimi, *supra* note 39.

51 Falliere, et al., *supra* note 44, at 3.

52 *See* Nat'l Research Council, Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities 118 (William A. Owens, Kenneth W. Dam, Herbert S. Lin eds., 2009).

In cyber operations parlance, intelligence collection and cyber reconnaissance or computer network exploitation (CNE) are often distinguished from CNA.[53] Analyzed independently, intelligence functions and CNE, such as those performed in support of Stuxnet, likely do not rise to the level of attack. Yet understood in context, many CNE could be understood as essential sub-components of an operation constituting an attack. CNE conducted immediately prior to an attack or even concurrently with an operation to damage or destroy property, such as appears to have been the case in the Stuxnet operation, present a strong case for satisfying logical and legal thresholds of attack. In law-of-war parlance, though not independently qualifying as attacks, CNE may nonetheless be said to constitute "direct participation in hostilities" – a function traditionally reserved to the combatant class. Thus questions arise concerning who might permissibly conduct CNE and even weapon design in support of CNA. Would the use of persons not meeting the four 1874 criteria described above warrant denial of combatant status and the consequences of combatancy? And would a State employing civilians to perform the intelligence functions, attack execution, or any of the other operations essential to a successful destructive CNA such as Stuxnet be in violation of the law of war?

The traditional and presently the majority answer is "yes." Respected international legal scholars have applied the 1874 criteria of combatant status to evaluate the question of lawful participation in cyber warfare. Nearly all conclude that only members of armed forces or organizations meeting the four 1874 criteria for combatant status should be employed to carry out CNA.[54] Most prescribe that States incorporate their cyber warriors into the regular armed forces or confer on them some military status. Few if any scholars or practitioners have deemed the 1874 inadequate or inapposite to the cyber context. Even scholars advocating innovative approaches to evaluating lawful participation in hostilities hew towards or even incorporate the four 1874 combatant status criteria.[55]

Yet, as I have suggested previously, several factors counsel skepticism towards unquestioning reliance on the 1874 criteria to evaluate combatant status in cyber warfare. First, States may already have heavily incorporated civilians into the agencies that support and conduct CNA on their behalf, making their direct participation in CNA likely if not certain. While the staffing details of States cyber war apparatuses are not publicly available, the executive mandates of several U.S. agencies suggest involvement in response to and use of CNA. In addition to the Department of Defense and its subordinate intelligence agencies (staffed in significant part

53 Computer network exploitation (CNE) refers to efforts to penetrate systems to gain information on the system and its vulnerabilities, thus acting as a tool for intelligence collection rather than system destruction. *See* Clay Wilson, *Information Operations, Electronic Warfare, and Cyberwar: Capabilities and Related Policy Issues 5* (Cong. Research Serv., CRS Report for Congress Order Code RL31787, Mar. 20, 2007), available at http://www.fas.org/sgp/crs/natsec/RL31787.pdf.

54 *See e.g.* Davis Brown, *A Proposal for an International Convention to Regulate the Use of Information Systems in Armed Conflict*, 47 Harvard International Law Journal 179, 187 (2006); Adam Sherman, *Forward unto the Digital Breach: Exploring the Legal Status of Tomorrow's High-Tech Warriors*, 5 Chicago Journal of International Law 335, 339–40 (2004); Louise Doswald-Beck, *Computer Network Attack and the International Law of Armed Conflict*, in 76 International Legal Studies: Computer Network Attack and International Law (Michael N. Schmitt & Brian T. O'Donnell eds., 2002)(concluding that rules guiding combatant classification and privilege should be no different in CNA); Michael N. Schmitt, *Wired Warfare: Computer Network Attack and the Jus in Bello*, in Computer Network Attack and International Law (Michael N. Schmitt & Brian T. O'Donnell eds., 2002)(concluding that civilians participating in CNAs that actually or foreseeably result in injury, death, damage, or destruction would be illegal combatants).

55 *See* Geoffrey S. Corn, *Unarmed but How Dangerous? Civilian Augmentees, the Law of Armed Conflict, and the Search for a More Effective Test for Permissible Civilian Battlefield Functions*, 2 National Security Law & Policy 257 (2008).

by civilian personnel), the U.S. Department of Homeland Security, the Central Intelligence Agency, and the Federal Bureau of Investigation share responsibility for defending against and responding to national security threats such as CNA.[56] Furthermore, the nature of both the physical and human capital of cyber warfare suggests a strong likelihood of significant civilian involvement. The programming expertise and vast network infrastructure of the civilian community and private sector make incorporation of their efforts into CNA seemingly irresistible.[57] Few, if any, of the actors holding the requisite expertise qualify as combatants under the 1874 criteria, thus rendering likely or even extant State practice inconsistent with presently conceived international law.

In addition to better aligning law and State practice, abandoning rote application of the 1874 combatant criteria accounts for their reduced practical relevance in cyberspace. First, although it is a significant indication of State affiliation or imprimatur, a criterion I will recommend be retained, the command criterion itself is a formalistic and empty requirement in cyber warfare. While the command requirement excludes individual actors and therefore preserves the collective nature of war, command remains essential in only a loose sense to cyberspace. Unlike their kinetic counterparts, cyber combatants are not typically isolated or removed from supervision or political leadership. The actions of cyber combatants seem susceptible to any number of management and supervision schemes including civilian or administrative oversight. In cyber warfare, requiring strict or formal military command is not uniquely suited to maintaining accountability or control of personnel carrying out CNA. If preserved as a prerequisite to combatant status in cyberspace, subordination to military command might easily be reduced to empty formalism – simply a paper drill conferring military status or bureaucratically incorporating what remains for all intent and purpose a civilian organization into an ersatz armed force of the State. Such hollow, *pro forma* measures would accomplish little, if anything, practically and would inevitably reduce respect for any law understood to require such steps.

The nature and circumstances of cyber warfare also undermine traditional application the second and third of the 1874 Brussels Declaration combatant status criteria. Because CNA constitute truly remote, over-the-horizon engagements, the classic requirements of distinctive insignia and carrying arms openly are of greatly reduced utility. Visually, cyber warriors are extremely unlikely to confront their foes. Unlike conventional kinetic attack, where attackers select targets on the basis of outward appearances or where defenders respond to the appearance of persons conducting the attack, CNA targets are selected on the basis of functionality or informational value. Far more than the outward appearance of individuals conducting CNA, distinction in CNA demands attention to the actual conduct of the attack – the target chosen, the pathways of entry, and the means used to achieve destruction or other harmful effects.

The final requirement of the 1874 criteria, that combatants' operations comply with the law of war enforceable through an internal disciplinary system retains much of its force but nonetheless takes on relatively reduced significance as well in CNA. While this fourth criterion undoubtedly

56 Exec. Order No. 12,333, 3 C.F.R. 200 (1982), *reprinted in* 50 U.S.C. § 401 app. at 542–51 (2006).

57 *See* Susan W. Brenner, Cyberthreats: The emerging Fault Lines of the Nation State (2009) (arguing for better integration of civilian law enforcement and intelligence organizations and military response to cyber attacks); Susan W. Brenner & Leo L. Clarke, *Civilians in Cyberwarfare: Conscripts*, 43 Vanderbilt Journal of Transnational Law 1011 (2010); Susan W. Brenner & Leo L. Clarke, Civilians in Cyberwarfare: Casualties, 13 Southern Methodist University Science & Technology Law Review 249 (2010).

retains its normative appeal and humanitarian effect with respect to observance of the law of war, the requirement of an internal disciplinary system one finds in later expressions of the fourth criterion seems largely inapposite to CNA. Envisioned as portable justice mechanisms capable of following armed forces wherever they operate and overcoming jurisdictional defects of their civilian counterparts, internal, military justice systems were a necessary corollary to command. Military justice was essential to enforcing discipline and preventing lawless exploitation of the battlefield by super-empowered belligerents. CNAs rarely, if ever, call for such jurisdictional portability and insularity. Participants in CNA need not be geographically displaced from civilian municipal justice systems. While investigating and prosecuting cyber war crimes would undoubtedly present great technical and legal challenges, the challenges specific to the kinetic battlefield deployment seem not to carry over in sufficient scale to warrant subjection to an internal military disciplinary system as a criterion for combatant status. In fact, the law of war increasingly forms part of States' domestic criminal codes, permitting meaningful civilian prosecution of war crimes committed by cyber warriors.[58] Finally, because senior leaders and legal advisors, presumably better-steeped in the law of war, can position themselves literally at arm's length from subordinate cyber combatants, the need for fourth criterion overall is perhaps reduced.

In contrast to the four 1874 Brussels Declaration criteria for combatant status, the single criterion of State affiliation far better supports the likely future of State practice in cyber warfare and vindicates the still important normative goals of the law of war. First, State affiliation preserves concern for the principle of distinction in CNA. If concern for distinction persists in cyber warfare, concern lies not so much with the identities and appearances of participants in CNA as much as with their weapons and the appearances generated by the attack itself. CNA have great capacity to confound their targets. Thus, the true challenge from CNA with respect to distinction may result not from civilian participation, rather from efforts to disguise the true source of the attack. CNA routed through civilian servers or programmed to appear as though they originated from civilian institutions may in fact run afoul of states' duty to bear arms openly in the attack. Exploration of this aspect of CNA's relation to distinction, however, is better left to a dedicated legal discussion of means and methods in CNA.

While considerable clarification of distinction in the context of CNA is required, state affiliation ensures that attacks remain subject to the existing international legal framework. In particular, the war crime of perfidy may present a more effective check against CNA exploiting peaceful or civilian networks as cover than restricting combatant status. Examining distinction, specifically the duty for those taking a direct part in hostilities to make themselves distinct from civilians, civilian CNA participants do not fail distinction by virtue of intentional perfidy. The intent of States' use of civilians in CNA is not to take advantage of enemy forbearance in targeting such civilians. More likely economic, training, and recruitment limitations drive the use of civilians in CNA. Situated far from the battlefield, if cyber warfare can be said to have a battlefield,[59] civilians participating in CNA do not present a confused picture to the enemy from the perspective of distinction. The likelihood that state-sponsored CNA could be misattributed

58 International Committee of the Red Cross, *International Humanitarian Law National Implementation Database, available at* http://www.icrc.org/ihl-nat.nsf/WebALL!OpenView (providing State-by-State information on domestic implementation of the law of war).

59 *See* Michael N. Schmitt, *The Principle of Discrimination in 21st Century Warfare*, 2 Yale Hum. Rts. & Dev. L.J. 143, 161–62 (1999). Battlespace describes both "virtual and non-linear loci of combat." *Id*. at 161.

to innocent civilian assets and systems make distinction of means far more important than distinction of personnel launching attacks.

In addition, reliance on state affiliation as the sole criterion for lawful participation in CNA presents no greater threat to discipline in warfare. While civilians participating in CNA are ordinarily not subject to internal military disciplinary systems, the increasing well-developed legal regimes that prosecute and punish war crimes operate nonetheless and vindicate concerns for discipline and humanity. As outlined above, when adopted by the 1949 Convention the criterion of exposure to an internal disciplinary system as a precondition to combatant status seemed reasonable. International enforcement bodies such as the International Criminal Court did not exist. Moreover, the international community's political will to convene *ad hoc* tribunals to prosecute war crimes appeared spotty and susceptible to victor's bias. Few if any international war crimes enjoyed domestic implementation or incorporation into states' domestic criminal codes. What enforcement of war crimes law existed was constrained largely to members of armed forces. The wide-scale incorporation of the law of war into domestic criminal mechanisms where civilians are equally susceptible to war crimes prosecution, including forms of vicarious liability, mitigates concerns that merely requiring State affiliation would inadequately serve the important concern of combatant discipline and humanity.

4. CONCLUSION

> *"The United States is actively engaged in the continuing development of norms of responsible state behavior in cyberspace, making clear that as a matter of U.S. policy, long-standing international norms guiding state behavior also apply equally in cyberspace. Among these, applying the tenets of the law of armed conflict are critical to this vision, although cyberspace's unique aspects may require clarifications in certain areas."*[60]

Like other innovations in warfare, cyber warfare will likely demand altered understandings of existing legal and operational precepts. While the principles and even many of the particulars of the law of war will in large part suffice to ensure a level of humanity and order in cyber warfare, to expect an unchanged or static legal convention to operate is unrealistic and would be ultimately self-defeating. As the above quotation makes clear, cyber hostilities will demand States issue clarifications and even operate under glosses on accepted tenets of the law of war.

The standards for combatant status in cyber warfare appear to be ripe for such a clarification. Recent developments including the Stuxnet attack make clear that executing successful CNA will place intense demands on States' human and technical capital, inducing many to resort to segments of their civilian population's expertise and infrastructure. Given the important consequences of determinations of combatant status, the extent to which the law of war accepts or condemns States' resort to their technical and personal capital may be one of the most important legal questions surrounding cyber warfare.

60 *Cyberspace Policy Report, supra* note 42, at 7-8.

Well-suited to the battlefields they imagined and those of over 100 years of succeeding armed conflicts, the 1874 Brussels Declaration combatant criteria continue to perform a useful sorting function on kinetic battlefields pitting visible adversaries against one another. The important principles of distinction and discipline draw direct support for each of the four criteria. Yet transposed to the realm of cyber warfare and use to evaluate the propriety of participation in hostilities by cyber warriors, the 1874 criteria appear dated and detached. Mainstream legal scholarship on combatancy in cyber warfare would exclude many cyber warriors from the class of lawful combatant unnecessarily and likely to the great disruption of existing or planned State practice while achieving little payout with respect to humanitarian ideals. Secondary rules, such as the proposed State affiliation gloss on the requirements of combatant status will both take account of emerging State practice while supporting the critically important notion captured in the primary rule of distinguishing combatants from civilians.

Idealized portraits of war are not entirely fatuous. Capturing our highest humanitarian aspirations in international law at once testifies to our shared interest in shielding the innocent and stricken from the horrors of war and reveals our belief in the power of law to work for good, even in the face of war. Yet alongside these aspirations must operate realistic and pragmatic understandings of the limits of combatants' capabilities and characteristics. Such understandings and interpretations secure law's voice in war and build the confidence in its end users necessary for its further development and efficacy.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Direct Participation in Cyber Hostilities: Terms of Reference for Like-Minded States?

Jody M. Prescott
West Point Center for the Rule of Law
West Point, New York, U.S.A.
jody.prescott@us.army.mil

Abstract: According to its recently published cyber strategy, the U.S. seeks to develop international consensus on how traditional law of armed conflict (LOAC) norms and understandings are modified and applied in cyberspace to help secure this global commons. Although the International Committee of the Red Cross's Interpretive Guidance on Direct Participation in Hostilities and the recent U.S. cyber strategy documents and policy statements are very different in many ways, examination of the relationships between their different aspects could be very useful in setting terms of reference framing the discussions which must occur to develop consensus on how LOAC rules and understandings regarding direct participation in hostilities could be adapted for use in cyberspace. This requires identification of their respective strengths and weaknesses, and potential areas of common ground between them. To be useful, this examination must include consideration of the significance of rules of engagement, formulations of hostile intent, and the proper inferences to be drawn from intelligence analyses as well as the legal standards by which direct participation in hostilities is determined.

Keywords: *direct participation, hostilities, cyber conflicts, law of armed conflict*

1. INTRODUCTION

The recently issued U.S. *International Strategy for Cyberspace* posits an end state in which cyberspace is "an open, interoperable, secure, and reliable information and communications infrastructure that supports international trade and commerce, strengthens international security, and fosters free expression and innovation."[1] To reach that goal, the U.S. foresees coordinated, international action as necessary to "build and sustain an environment in which norms of responsible behavior guide states' actions, sustain partnerships, and support the rule of law in cyberspace."[2] This end state would be fostered by norms resulting from the U.S.'s "work with like-minded states to establish an environment of expectations [...] that ground foreign

1 The White House, International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World, 8 (May 2011), *available at* http://www.whitehouse.gov/sites/default/files/rss_viewer/international_strategy_for_cyberspace.pdf [hereinafter "International Strategy"].

2 *Id.*

and defense policies and guide international partnerships."[3] Working with "like-minded" states is important to the U.S. because it believes the current unsettled state of cyberspace has resulted in part from "governments seeking to exercise traditional national power through cyberspace" without "clearly agreed-upon norms for acceptable state behavior."[4] In addressing this situation, the U.S. believes that "[l]ong-standing international norms guiding state behavior – in times of peace and conflict – also apply in cyberspace," but that the "unique attributes of networked technology require additional work to clarify how these norms apply and what additional understandings might be necessary to supplement them."[5]

This paper suggests that a comparison of the INTERPRETIVE GUIDANCE[6] of the International Committee of the Red Cross (ICRC) on direct participation in hostilities could, in conjunction with the *International Strategy* and subsequent U.S. Department of Defense (DoD) cyber strategy documents and policy statements, help set terms of reference to frame the discussions concerning the application of the principle of direct participation in hostilities in cyberspace. This requires, however, a frank assessment of the conceptual weaknesses and strengths of each approach, where they differ, and where there may be common ground. Thus, this paper will first set out the main points of the INTERPRETIVE GUIDANCE, particularly noting its consideration of cyber conflict. Next, it will examine the shortcomings in the INTERPRETIVE GUIDANCE's approach to direct participation in modern armed conflicts. Against this backdrop, the apparent U.S. position will be examined to identify possible trends in the development of concepts related to direct participation in hostilities, and the ramifications of these trends were they to become operationalized. In conclusion, this paper will suggest that although the development of consensus among the "like-minded" on the topic of direct participation in hostilities will not likely be simple nor will it be smooth, its progress would be furthered by an understanding of how the relationships between the differences and the similarities in the ICRC and U.S. positions help set terms of reference for the discussions that must occur.

2. THE INTERPRETIVE GUIDANCE

The INTERPRETIVE GUIDANCE sets out three cumulative elements that must be met before an individual is deemed to have lost the presumption in favor of finding him to be a protected civilian in both international and non-international armed conflict: a threshold of harm, direct causation, and a belligerent nexus.

A. Threshold of Harm

As to the threshold of harm, the INTERPRETIVE GUIDANCE notes that if the reasonable result of an act would be "harm of a specifically *military nature*," this requirement would generally be met "regardless of the quantitative gravity" of the adverse effect.[7] As an example,

3 *Id*. at 9.
4 *Id*.
5 *Id*.
6 NILS MELZER, INT'L COMM. OF THE RED CROSS, INTERPRETIVE GUIDANCE ON THE NOTION OF DIRECT PARTICIPATION UNDER INTERNATIONAL HUMANITARIAN LAW 20 (2009), *available at* http://www.icrc.org/eng/assets/files/other/icrc-002-0990.pdf [hereinafter "INTERPRETIVE GUIDANCE"]. It was compiled on the basis of reports generated from meetings of international experts in the law of armed conflict (LOAC) held between 2003 and 2008. Id. at 8.
7 *Id*. at 47.

"electronic interference with military computer networks could […] suffice, whether through computer network attacks […] or computer network exploitation."[8] However, were the harm not military, the "specific act must be likely to cause at least death [or] injury, or destruction" of property.[9] Accordingly, although acts such as "the manipulation of computer networks [might] have a serious impact on public security, health and commerce," this impact itself would be insufficient to cross the threshold of harm.[10]

Some writers suggest that such a standard would be too restrictive, and that consistent with article 51.2 of Additional Protocol I[11] (prohibiting measures that terrorize civilian populations), injury should include "severe physical or mental suffering."[12] Further, the "loss of intangible assets (e.g., funds held electronically in a banking system) that are directly transformable into tangible assets (e.g., currency or purchasable objects) could be" within the definition of property.[13] The INTERPRETIVE GUIDANCE, however, focuses on harm that occurs in the geophysical world as a result of physical violence.[14]

B. Direct Causation

The INTERPRETIVE GUIDANCE notes that in keeping with the distinction set out in LOAC between direct participation in hostilities that would render an ordinarily protected civilian targetable and indirect participation (such as working in a munitions factory) which would not remove that protection, the difference between the two must "correspond […] to that between direct and indirect causation of harm."[15] Accordingly, "[i]n the present context, direct causation should be understood as meaning that the harm […] must be brought about in one causal step."[16] Examples of actions that would not meet this standard include capacity building through recruiting and training personnel.[17] The INTERPRETIVE GUIDANCE notes that not all of the experts agreed to this formulation, citing examples such as the building of improvised explosive devices (IEDs) and missiles by non-state actors as being more than "mere capacity building […] and becom[ing] measures preparatory to a concrete military operation."[18] As to the timeframe during which direct participation in hostilities exists, the INTERPRETIVE GUIDANCE states that actions in preparation for an "act of direct participation in hostilities, as well as deployment to and return from the location of its execution, constitute an integral part of that attack."[19] If, however,

> "the execution of a hostile act does not require geographic displacement, as may be the case with computer network attacks[,] the duration of direct participation in hostilities will

8 *Id*. at 48.
9 *Id*.
10 *Id*. at 50.
11 Protocol Additional to the Geneva Conventions of Aug. 12, 1949, and relating to the Protection of Victims of International Armed Conflicts, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter "AP I"].
12 Michael N. Schmitt, Heather A. Harrison & Thomas C. Wingfield, *Computers and War: The Legal Battlespace*, Background Paper prepared for Informal High Level Expert Meeting on Current Challenges to International Humanitarian Law, Cambridge, June 25-27, 5 (2004).
13 *Id*.
14 INTERPRETIVE GUIDANCE, supra note 6, at 20, 49-50.
15 *Id*. at 52.
16 *Id*. at 53.
17 *Id*. at 54.
18 *Id*. at 54 n.125.
19 *Id*. at 65. These acts must be of "a specific military nature and so closely linked to the subsequent execution of a specific hostile act that they already constitute an integral part of that attack." Id. at 65-66.

be restricted to the immediate execution of the act and preparatory measures forming an integral part of that attack."[20]

C. Belligerent Nexus

As to the third element, the purpose of the act being to directly cause an effect which crosses the required threshold of harm, the INTERPRETIVE GUIDANCE states that before an act could be considered direct participation, it must "be objectively likely to inflict harm that meets the first two criteria [and] specifically designed *to do so in support of a party to an armed conflict and to the detriment of another*."[21] The INTERPRETIVE GUIDANCE holds that such a group must belong to a party to the conflict; a status which "can be shown by conclusive behavior that makes it clear for which party the group is fighting."[22] As Professor Michael Schmitt has noted, this "would exclude those organized armed groups in an international armed conflict that might be directing cyber attacks against one of the parties for reasons other than support of the opposing party," such as unaffiliated patriotic hacker groups.[23]

The INTERPRETIVE GUIDANCE notes that not all uses of armed force in an armed conflict will necessarily be considered part of the on-going hostilities. For example, quelling civil unrest which is unrelated to the actual fighting in a combat zone would be excluded,[24] and armed forces engaged in such activities would find their use of force restricted to applications consistent with law enforcement standards and concepts of individual self-defense.[25] However, it also notes that in many armed conflicts, serious criminals may operate such that "it is difficult to distinguish hostilities from violent crime unrelated to, or merely facilitated by, the armed conflict."[26] In light of the increasing incidence of cybercrime, distinguishing between cyberspace actors who are directly participating in a conflict and those who are merely opportunistic criminals could prove even more challenging than in the geophysical world.

D. Continuous Combat Function

The INTERPRETIVE GUIDANCE also sets out the concept of "continuous combat function,"[27] by which individuals whose functions as part of organized non-state actor armed forces "involve […] the preparation, execution, or command of acts or operations amounting to direct participation in hostilities" may be targeted even if not actively participating in hostilities at the time they are engaged.[28] This is intended to distinguish them from "civilians who participate in hostilities on a merely spontaneous, sporadic, or unorganized basis, or who assume exclusively political, administrative or other non-combat functions."[29] This latter category of individuals

20 *Id*. at 68.
21 *Id*. at 58 (emphasis in original).
22 *Id*. at 35.
23 Michael N. Schmitt, *Cyber Operations and the Jus in Bello: Key Issues*, 87 INT'L LAW STUDIES, INTERNATIONAL LAW AND THE CHANGING CHARACTER OF WAR, 89, 100 (Raul A. Pedrozo & Daria P. Wollschlaeger eds. 2011) [hereinafter "*Cyber Operations*"].
24 INTERPRETIVE GUIDANCE, *supra* note 6, at 62-63.
25 *Id*. at 76.
26 *Id*. at 68.
27 *Id*. at 33.
28 *Id*. at 34.
29 *Id*.

could only be targeted for such time as they were taking a direct part in hostilities, as defined *supra*.[30]

The INTERPRETIVE GUIDANCE qualifies continuous combat function quite restrictively. First, for an individual to have membership in organized non-state actor armed forces, that person must assume a role that "corresponds to that collectively exercised by the group as a whole, namely, the conduct of hostilities on behalf of a non-state party to the conflict."[31] Second, the acts the individual commits in such a role must occur "in circumstances indicating that such conduct constitutes a continuous function rather than a spontaneous, sporadic, or temporary role assumed for the duration of a particular operation."[32]

The significance of the *group* purpose is fundamental to this concept, for as Professor Schmitt has noted, "the concept of armed forces makes no sense in the absence of a group purpose of violence."[33] Such a group could include "an on-line group [that has] a defined command structure and coordinate[s] its war-like activities" in cyberspace.[34] In Professor Schmitt's view, a group without a violent purpose "is but a collection of civilians", and its members only become targetable to the extent that their individual activities constitute direct participation in hostilities.[35] As a practical matter, however, given the fluid nature of identity in cyberspace, if intelligence showed that an individual member of such a group was directly participating in hostilities, and that similar groups ordinarily disguised their true purpose in part by vectoring war-like acts through a single member, it might be reasonably concluded that the requisite group purpose existed.

3. SHORTCOMINGS IN THE INTERPRETIVE GUIDANCE

A. The Standard of Decision

The first of the INTERPRETIVE GUIDANCE's four shortcomings lies in not following through to the logical conclusion that flows from its acknowledgment of the practical and situation-dependent standard to be used to determine whether an individual is a legitimate military target rather than a civilian. It notes that "all feasible precautions must be taken" to ensure that individuals who are targeted are in fact legitimate military targets, and not protected civilians. "[F]easible precautions" are "those which are practicable or practically possible taking into account all circumstances ruling at the time, including humanitarian and military considerations."[36] Accordingly, the INTERPRETIVE GUIDANCE notes that the standard of doubt to be applied in targeting decisions is not the same as that applied in criminal proceedings, and instead "must reflect the level of certainty that can reasonably be achieved

30 "Civilians lose protection against direct attack for the duration of each specific act amounting to direct participation in hostilities, whereas members of organized armed groups belonging to a non-state party to an armed conflict cease to be civilians[,] and lose protection against direct attack for as long as they assume their continuous combat function." *Id*. at 70.

31 *Id*. at 33.

32 *Id*.

33 *Cyber Operations, supra* note 23, at 99.

34 *Id*. at 98-99.

35 *Id*. at 99.

36 Final Report on the Meaning of Armed Conflict in International Law, Use of Force Committee, International Law Association, The Hague Conference, 75 (2010), *available at* http://www.ila-hq.org/en/publications/index.cfm.

in the circumstances."[37] The targeting decision must therefore consider factors such as "the intelligence available to the decision maker, the urgency of the situation, and the harm likely to result to the operating forces or to persons and objects protected against direct attack from an erroneous decision."[38]

These realities mean that the standard that is applied throughout the targeting process is in effect reasonable certainty under the circumstances.[39] Reasonable inferences will be developed as a result of continuing analysis of an incomplete and evolving intelligence picture, and the standard is therefore weighted towards providing significant latitude in the evaluation of the factors that establish direct participation in hostilities, and allowing action in response. Operationally, this reality tends to undermine the cumulative restrictions set out in the INTERPRETIVE GUIDANCE.

B. Dismissal of Hostile Intent

The second problem with the INTERPRETIVE GUIDANCE lies in its assessment of the concept of hostile intent as being too bound up with rules of engagement (ROE)[40] to be useful in determining the legal contours of direct participation in hostilities. Because the meeting of experts viewed hostile intent as a technical ROE term, and ROE as national political and command guidance on the use of armed force that did "not necessarily reflect the precise content of IHL", it was therefore "generally regarded as unhelpful, confusing or even dangerous to refer to hostile intent for the purpose of defining direct participation in hostilities."[41] However, the definition of hostile intent is completely relevant to a discussion of the definition of direct participation in cyber hostilities, because in many ways it sets the lowest threshold for activity that can be seen as justifying a lethal response from an opposing armed force in armed conflict involving unfriendly actors who do not necessarily identify themselves as being members of an organized armed force.

NATO ROE recognize that the different NATO member nations will have different interpretations of the right to engage in self-defense,[42] and to cross-level these inconsistencies ROE are provided for mission accomplishment that include the authority to respond to manifestations of hostile intent.[43] For example, NATO ROE Serial 421 provides that "[a]ttack against [designated] force(s) or [designated] target(s) demonstrating hostile intent (not constituting an imminent attack) against NATO/NATO-led forces is authorized."[44] The NATO ROE define hostile intent as having two elements: the "capability and preparedness of individuals, groups of personnel or units which pose a threat to inflict damage," and "evidence, including intelligence, which indicates an intention to attack or otherwise inflict damage."[45]

37 INTERPRETIVE GUIDANCE, *supra* note 6, at 76.
38 *Id*.
39 Joint Targeting Cycle and Collateral Damage Estimation Methodology (CDM), Briefing by DoD General Counsel, 26 (Nov. 10, 2009), *available at* http://www.nefafoundation.org/newsite/file/awlaki_DODUAVstrikes.pdf.
40 NATO defines ROE as "directives to military forces (including individuals) that define the circumstances, conditions, degree, and manner in which force, or actions which might be construed as provocative, may be applied." NORTH ATLANTIC TREATY ORGANIZATION, MILITARY COMMITTEE, MC 362/1, NATO RULES OF ENGAGEMENT, MC 362/1, 2 (June 30, 2003) [hereinafter "NATO ROE"].
41 INTERPRETIVE GUIDANCE, *supra* note 6, at 59 n.151.
42 NATO ROE, *supra* note 40, at 3-4.
43 *Id*. at ¶2, App. 1, Annex A.
44 *Id*. at A-19.
45 *Id*. at ¶3, App. 1, Annex A.

In illustrating this definition, the NATO ROE look in part to objective, physical indicators of ill intent, such as "manoeuvering into weapons launch positions," and non-tactical events such as the "increased movements of ammunition and the requisition of transport."[46] This definition also sets a threshold of harm to be used to help determine whether hostile intent is present, noting that "[i]solated acts of harassment, without intelligence or other information indicating an intention to attack or otherwise inflict damage, will not normally be considered hostile intent."[47]

The anonymity of cyber space, and the ability of unfriendly actors to "spoof" their true identities,[48] challenges the application of the principle of distinction to cyber actors. In those cases where the accurate identification of the cyber actor would be required before undertaking a certain response in the geophysical world, such as imposing economic sanctions or engaging the known digital infrastructure of a nation because its armed forces had apparently launched a cyber attack by proxy, attribution is of course a crucial issue. In the context of assessing whether an actor with an unknown identity is taking a direct part in hostilities as measured by an assessment of whether their intent is hostile, however, attribution to a particular state or non-state actor may not be necessary before engaging the threat.

C. Inaccurate View of the Intelligence Picture

The INTERPRETIVE GUIDANCE's third flaw is its inaccurate assumption of what targeting intelligence looks like, and its lack of discussion as to how reasonable inferences can be drawn from analyzing patterns of information that will work to fill in the gaps between actual data hard points. These inferences lend themselves to resolving doubt as to whether an individual is taking a direct part in hostilities without triggering the presumption of protected status, under the standard of reasonable certainty discussed *supra*. Although targeting intelligence may often be uneven in quality and depth, the INTERPRETIVE GUIDANCE appears to assume a very broad intelligence picture being available to militaries, one which is very detailed and capable of informing commanders and soldiers at various levels of the information they would need to make the informed decisions to comply with its recommendations. For example, in the determination of whether civilians meet the belligerent nexus element, it makes clear that it is not recommending assessing the subjective intent of the actor. However, it then provides the confusing example of civilians who might be unaware of the role they are playing in hostilities, by unknowingly transporting weapons for example. In this case, it states

> "[t]hey remain protected against direct attack despite the belligerent nexus of the military operation in which they are being instrumentalised. As a result, these civilians would have to be taken into account in the proportionality assessment during any military operation likely to inflict incidental harm on them."[49]

The chances of a targeting authority knowing that an individual transporting such a cargo was unaware of it are highly unlikely. The practical uselessness of this concept is demonstrated by the very fine distinction it attempts to draw between those who are executing a continuous combat function versus those whose war-like acts are "sporadic" or "spontaneous":

46 *Id*. at ¶4, App. 1, Annex A.
47 *Id*.
48 Jody Prescott, *War By Analogy: US Cyberspace Strategy And International Humanitarian Law*, 156 RUSI J. 32, 33-34 (Dec. 2010).
49 INTERPRETIVE GUIDANCE, *supra* note 6, at 60.

> "Where civilians engage in hostile acts on a persistently recurrent basis, it may be tempting to regard not only each hostile act as direct participation in hostilities, but even their continued intent to carry out unspecified hostile acts in the future. However, any extension of the concept of direct participation in hostilities beyond specific acts would blur the distinction made in IHL between temporary, activity-based loss of protection (due to direct participation in hostilities), and continuous, status or function-based loss of protection [...]."[50]

The INTERPRETIVE GUIDANCE provides no guidance that would help distinguish between reports of a series of war-like acts by an individual which are merely spontaneous as compared to reports on a person who commits the exact same sorts of acts but is exercising a continuous combat function. Instead, it posits that it is not operationally possible to "determine with a sufficient degree of reliability whether civilians not currently preparing or executing a hostile act have previously done so on a persistently recurrent basis and whether they have the continued intent to do it again."[51] Hypothetically, whether an individual has committed war-like acts in the past could be tracked by modern intelligence assets, if that information has been collected.[52] Communications intercepts or similar reports could indicate whether this person is participating in the planning of future war-like act. If "the principle of distinction must be applied based on information which is practically available and can reasonably be regarded as reliable in the prevailing circumstances,"[53] then the reasonable inferences that could be drawn from the information in this hypothetical would support an assessment of continuous combat function, rather than war-like spontaneity, on the part of the individual.

D. Too Restrictive Window of Direct Participation

The fourth shortcoming of the INTERPRETIVE GUIDANCE is its overly restrictive definition of the time frame within which those directly participating in cyber hostilities may be targeted. Restricting this attack window to just before, during, and immediately after a cyber event is at odds with the manner in which potential cyber attacks could occur. First, the nature of so-called "Zero Day"[54] defects in digital infrastructure means an unfriendly intrusion could evolve into a potentially catastrophic attack at near light-speed.[55] Second, at the moment it occurs, it is likely very challenging to quickly determine whether the intruder is an opposing state, a terrorist group, a cyber criminal, or a hacker.[56] Execution of a cyber attack might follow immediately after an intrusion, and the preparatory measures might either be invisible to the affected state or seem innocuous.[57] In Professor Schmitt's view, this means that "there may be no 'deployment' at all," since "only a computer, and not proximity to the target is required to

50 *Id*. at 45.
51 *Id*.
52 *See* Major General Michael T. Flynn, Captain Matt Pottinger & Paul D. Batchelor, *Fixing Intel: A Blueprint for Making Intelligence Relevant in Afghanistan, Voices from the Field*, CENTER FOR A NEW AMERICAN SECURITY, 7-8 (2010) (intelligence collection in Afghanistan focused on insurgent activity and identity).
53 INTERPRETIVE GUIDANCE, *supra* note 6, at 35.
54 William Jackson, *Malicious PDFs Exploit Zero-Day Vulnerability and Adobe Reader*, GOV'T COMPUTER NEWS, Feb. 20, 2009, *available at* http://gcn.com/articles/2009/02/20/pdf-zero-day-eploit.aspx.
55 *Cyber Operations, supra* note 23, at 102.
56 Committee on Offensive Information Warfare, *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities*, 18, 21, 90-91,William A. Owens, Kenneth W. Dam & Herbert S. Lin eds, NATIONAL RESEARCH COUNCIL (2009).
57 *Id*. at 90-93.

mount the operations."[58] Further, the cyber attack itself "might last only minutes, perhaps even seconds."[59] The Interpretive Guidance's restriction of direct participation in hostilities to the time of execution and just before or after would therefore "effectively extinguish the right to strike at direct participants."[60]

4. THE APPARENT U.S. PERSPECTIVE ON DIRECT PARTICIPATION

There is no single unclassified U.S. strategy document or policy statement that explicitly sets out how the U.S. understands and intends to apply the concept of direct participation in hostilities to cyber conflicts. Therefore, different unclassified strategy documents and policy statements must both be considered together and individually scrutinized to glean indications of how U.S. policy and thinking might be evolving in this regard. One fundamental theme runs through all the sources of the U.S. position, however: "cyberspace activities can have effects beyond networks; [and] such events may require responses in self-defense" and trigger "commitments [it has] with [its] military treaty partners [...]."[61]

A. The DoD Strategy for Operating in Cyberspace

Rather than focusing on the use of force, the unclassified version of the *DoD Strategy for Operating in Cyberspace* (DoD Strategy),[62] released two months after the publication of the International Strategy, instead describes complementary strategic initiatives which emphasize the need to create a well organized, trained and equipped cyber force structure; to develop partnerships with civilian governmental agencies, private industry, allies and other international partners; and the need to develop a national wellspring of talent and innovation to keep the U.S. military and industry competitive in the cyber arena. Although the DOD Strategy sets out the use of "active cyber defense" as an operating concept, it defines it in a fairly benign manner as the "synchronized, real-time capability to discover, detect, analyze and mitigate threats and vulnerabilities."[63]

To put the *DoD Strategy* into its proper perspective, however, it is useful to examine the statements made by U.S. officials regarding DoD's cyber strategy in general. First, the definition of "active cyber defense" in the *DoD Strategy* is not completely consistent with earlier statements made by U.S. officials that suggested that "active cyber defense" included operations within other nations' digital infrastructures.[64] Similarly, U.S. Deputy Secretary of Defense William J. Lynn remarked at the time the *DoD Strategy* was published that although he believed "destructive or disruptive cyber attacks that could have an impact *analogous* to physical hostilities" would

58 *Cyber Operations, supra* note 23, at 102.
59 *Id.*
60 *Id.*
61 *International Strategy, supra* note 1, at 11-14. NATO handles cyber incidents under the consultative procedures of Article IV of the NATO Treaty rather than as attacks under Article V. *NATO Agrees Common Approach to Cyber Defence*, EURACTIVE.COM, Apr. 4, 2008, *available at* http://www.euractiv.com/infosociety/nato-agrees-common-approach-cyber-defence/article-171377.
62 Department of Defense Strategy for Operating in Cyberspace, DOD, July 2010, *available at* http://www.defense.gov/news/d20110714cyber.pdf [hereinafter "DOD Strategy"].
63 *Id.* at 7.
64 Ellen Nakashima, *Pentagon considers preemptive strikes as part of cyber-defense strategy*, WASHINGTONPOST.COM, Aug. 28, 2010, available at http://www.washingtonpost.com/wp-dyn/content/article/2010/08/28/AR2010082803849_pf.html.

occur in the future, that "the vast majority of malicious cyber activity today d[id] not cross this threshold."[65] Deputy Secretary Lynn's use of the word "analogous" to describe the relationship between war-like acts in the geophysical world and significant ill-intended acts in cyberspace was likely deliberate, and it suggests that the classified version of the *DoD Strategy* does not reflect direct translation into it of LOAC rules and concepts applicable in the geophysical world. Prior to the *DoD Strategy*'s launch, statements by DoD officials had indicated instead that it would be based on a concept of "equivalence" between geophysical world hostilities and unfriendly acts in cyberspace to guide its use of force in the latter domain.[66] On the spectrum of similarity, "equivalence" would suggest a more literal adoption of LOAC concepts and applications than would "analogy".

B. The DoD Cyber Policy Report

In November 2011, DoD provided the U.S. Congress with a report on the status of DoD's efforts to operationalize LOAC concepts in cyberspace.[67] The *Cyber Report* recognized the importance of establishing the identity of unfriendly actors, because cyberspace's "unique characteristics [could] make the danger of escalation especially acute. For instance, the speed of action and dynamism inherent in cyberspace, challenges of anonymity, and widespread availability of malicious tools can compound communications and increase opportunities for misinterpretation."[68] It noted DoD's work "with international partners to bolster cyber forensics capabilities," and very intriguingly, its efforts to "assess the identity of [an] attacker via behavior-based algorithms."[69] Complementing these efforts, the *Cyber Report* noted DoD's intent "to expand and deploy applications that detect, track and report malicious activities across all DoD networks and information systems on a near real-time basis."[70]

The *Cyber Report* also described the scope of the challenge confronting intelligence specialists, noting that "[t]he often low cost of developing malicious code and the high number and variety of actors in cyberspace make the discovery and tracking of malicious cyber tools difficult."[71] Further, "most of the technology used in this context is inherently dual-use, and even software might be minimally repurposed for malicious action,"[72] which made it even more difficult to definitively recognize and effectively track unfriendly cyber actors. Despite these difficulties, it stated that as with military intelligence operations in general, cyber intelligence operations were "governed by long-standing and well-established considerations."[73] However, perhaps in an implicit nod to an aggressive theory of active cyber defense, the report noted "the possibility that those operations could be considered a hostile act."[74]

65 William J. Lynn, Remarks on the Department of Defense Cyber Strategy, speech made in Washington, D.C. (July 14, 2011), *available at* http://www.defense.gov/Speeches/Speech.aspx?SpeechID=1593 (emphasis added).

66 Siobhan Gorman & Julian E. Barnes, Cyber Combat: *Act of War*, WSJ.COM, May 31, 2011, *available at* http://online.wsj.com/article/SB10001424052702304563104576355623135782718.html.

67 DOD Cyber Policy Report Pursuant to Section 934 of the NDAA of FY 2011(Nov. 2011), *available at* http://www.defense.gov/home/features/2011/0411_cyberstrategy/docs/NDAA%20Section%20934%20Report_For%20webpage.pdf [hereinafter "Cyber Report"].

68 *Id*. at 5.

69 *Id*.

70 *Id*.

71 *Id*. at 8.

72 *Id*.

73 *Id*. at 7.

74 *Id*.

Regarding cyber ROE, the *Cyber Report* stated that response options available to the President "may include using cyber and/or kinetic capabilities,"[75] which means that any potential attacker of U.S. cyberspace interests must consider not just the possibility and risk of a U.S. cyber response, but also the possibility of individuals and units conducting the attack and their equipment being engaged in the geophysical world. The *Cyber Report* also stated that the U.S. cyber ROE reflect "the interconnectedness and the speed that defines cyberspace," and that therefore they "reflect: the implications of cyber threats; the operational demands of DoD's continuous, world-wide operations; and the need to minimize disruption from collateral effects on networked infrastructure."[76] Further, the *Cyber Report* noted that "[a]s in the physical world, a determination of what is a 'threat or use of force' in cyberspace must be made in the context in which the activity occurs, and it involves an analysis by the affected states of the effect and purpose of the actions in question."[77] Together, these statements emphasize the crucial importance of the internet to U.S. military operations, and suggest that the cyber ROE provide significant latitude to engage on the basis of hostile intent or hostile act.

The *Cyber Report* suggests that DoD is in fact operationalizing LOAC concepts in cyberspace in an "analogous" rather than an "equivalent" fashion. In general, it notes that "DoD will conduct offensive cyber operations in a manner consistent with the policy principles and legal regimes that the department follows for kinetic capabilities, including the law of armed conflict."[78] Importantly, this consistency is at high and abstract level, and consistency is itself a lesser state of compliance than conformance. The *Cyber Report*'s treatment of the issue of potential violations of third nations' sovereignty rights also suggests this. The *Cyber Report* states that in the case of a neutral third country finding itself involved in a cyber threat to the U.S., DoD would adhere to LOAC principles"[79], and that DoD's responses could "include taking actions short of the use of force as understood in international law."[80] However, a number of factors would need to be considered in each case, including the "[n]ature of the act, [the] role of the 3rd country, its ability and willingness to respond effectively, and potential issues of sovereignty."[81]

5. POTENTIAL RAMIFICATIONS OF THE U.S. CYBER STRATEGY

A. War by Analogy

If cyber conflict is seen as only analogous to war in the geophysical world, then the translation of geophysical LOAC rules and interpretations into cyber LOAC norms and understandings will likely reflect this perspective. If the assessment of the U.S. position *supra* is correct, then the U.S. application of this perspective regarding LOAC seems to be the inclusion of LOAC principles and rules as factors to be considered in whether to take action, along with very functional concerns of practical impact on U.S. interests. This approach presents two potential problems, the first of which is whether the U.S. would be able to persuade a coalition of the like-minded of sufficient international stature to not just agree to this approach, but to the

75 *Id*. at 4.
76 *Id*. at 6.
77 *Id*. at 9.
78 *Id*. at 5.
79 *Id*. at 8.
80 *Id*.
81 *Id*.

specific factors to be considered and any weighting of them in the decision making that would be required as well. Second, given that the U.S. reserves the right to respond to unfriendly cyber action by kinetic action in the geophysical world, and that actions in cyberspace could conceivably ripple into the geophysical world as well, conducting cyber war could become like a game of three-dimensional chess, with different rules on different levels.[82] This would require commanders and legal advisors to not just be familiar with the effects of technology in cyberspace and how the agreed-upon analogous norms applied; they would also need to be able to simultaneously track the effects and the traditional LOAC rules applicable to those effects in the geophysical world. The training, educational and experiential requirements that would need to be met by the individuals filling these positions, to say nothing of the doctrine and educational infrastructure that would need to be built to produce such soldiers, would require a significant investment by nations to create these capabilities.

B. Cyber Due Diligence

The *International Strategy* describes "cyber due diligence" as an emerging norm essential to cyberspace's proper use. This term is defined as states' obligations to protect their "information infrastructures and secure national systems from damage or misuse."[83] As noted *supra*, the *DoD Strategy* is based in part on the employment of "new defense operating concepts to protect DoD networks and systems," and this includes measures to better train DoD personnel and hold them accountable for the proper secure use of digital infrastructure and to prevent intrusions from occurring.[84]

Neutral states are required under international law to enforce their neutrality and prevent parties in armed conflicts from using their territories as bases from which the parties could launch attacks against one another. If a state does not protect its neutrality, whether through lack of will or capacity, it risks being seen by the party receiving attacks from its territory as a co-belligerent. The attacked party might then engage its attackers on the sovereign territory of the ostensibly neutral nation, and in that fashion the neutral nation finds that it has become a direct participant in the conflict.[85] As noted *supra*, the *Cyber Report* sets out a list of factors that would be considered in deciding whether to engage a cyber threat located in a third country, and whether the country is exercising cyber due diligence is arguably included within the factor of whether the country has the capability and willingness to deal with the threat effectively itself. Sovereignty as a consideration is expressed in terms of how the U.S. might handle potential sovereignty issues, which is a functional calculus quite different than the third country's sovereignty itself being a factor. The concept of cyber due diligence, therefore, may have the effect of expanding the concept of direct participation in hostilities through loosening the restrictions on infringing upon another nation's cyber sovereignty.

C. Hostile Intent and Hostile Acts

The U.S. Standing ROE allow its forces to respond with lethal force to acts they perceive to be hostile. "Hostile acts" are defined broadly as "attack[s] or other use[s] of force against

82 Prescott, *supra* note 48, at 35.
83 *International Strategy, supra* note 1, at 10.
84 *DoD Strategy, supra* note 62, at 7.
85 Tess Bridgeman, *The Law of Neutrality and the Conflict with Al Qaeda*, 85 N.Y.U. L. REV. 1186, 1200 n.75 (2010).

the [U.S.], U.S. Forces, or other designated persons or property."[86] The examples provided to illustrate the scope of acts considered hostile confirm this broad application, and "include[s] force used directly to preclude or impede the mission and/or duties of U.S. personnel or vital [U.S. government] property." "Hostile intent" is defined just as broadly,[87] and both U.S. definitions are less restrictive than their NATO ROE counterparts.[88]

Examination of the U.S. position suggests that U.S. cyber ROE provide significant latitude to engage perceived cyber threats. The *Cyber Report* appears to premise action in cyberspace largely upon perception of hostile intent, expressed or implied, and hostile acts.[89] Presumably, because of the speed with which cyber weapons could be deployed, relying only upon cyber due diligence presents too great a risk of intrusion by unfriendly actors into DoD networks. Determining whether an actor is demonstrating hostile intent may require cyber operators to conduct searches for certain malicious code in targeted software, regardless of where in the geophysical world those programs actually resided, as part of active cyber defense.[90] Thus, hostile intent might be deduced from a characteristic of malware's composition without it actually being employed. Interestingly, the U.S. appears to realize that such actions on its part could be perceived as hostile acts, which suggests that the U.S. could, were similar actions undertaken within its digital infrastructure, view them the same way.

D. Threshold of Harm

Although the INTERPRETIVE GUIDANCE appears to set a threshold of harm caused by action against military assets and capabilities lower than the U.S. position's, this may actually be an area of common ground between the two positions. The INTERPRETIVE GUIDANCE notes that if the reasonable result of an act would be "harm of a specifically *military nature*," the threshold of harm requirement would generally be met "regardless of the quantitative gravity" of the adverse effect.[91] The *Cyber Report*, however, states only that hostile acts must be significant to be actionable.[92]

Professor Nils Melzer notes that "it could be argued that cyber attacks unlikely to result in death, injury or destruction could still amount to an 'armed attack' if they aim to incapacitate 'critical infrastructures' within the sphere of sovereignty of another state."[93] In the absence of military harm, however, it is not clear that such actions would result in their perpetrators being targetable if the "attack" resulted in no observable destruction in the geophysical world.[94] The U.S., however, is apparently taking an assessment of effects approach to making such a determination across the board. Presumably, this means guidelines as to significance would

86 INSTRUCTION 3121.01B, STANDING RULES OF ENGAGEMENT/STANDING RULES FOR THE USE OF FORCE FOR U.S. FORCES, CHAIRMAN OF THE JOINT CHIEFS OF STAFF, ¶e, A-3, Enclosure A (Jun. 13, 2005).

87 *Id*. at ¶f, A-3, Enclosure A.

88 *See* NATO ROE, *supra* note 40, at ¶¶3-5, App.1, Annex 1.

89 *Cyber Report, supra* note 67, at 3-4, 6.

90 In response to a question whether the U.S. would be able to prevent a cyber attack before it registered in the U.S., General Alexander has testified before the U.S. Congress that he is seeking ROE "to protect and prevent" cyber attack. Shaun Waterman, *Cyberwarfare rules still being written*, WASHINGTONTIMES.COM, *available at* http://www.washingtontimes.com/news/2012/mar/20/cyberwarfare-rules-still-being-written/.

91 INTERPRETIVE GUIDANCE, *supra* note 5, at 47.

92 *Cyber Report, supra* note 67, at 4.

93 Nils Melzer, *Cyber Warfare and International Law*, UNIDIR Resources, 14-16 (2011), *available at* http://www.unidir.org/pdf/activites/pdf2-act649.pdf.

94 *Id*. at 28, 31.

be consulted in each case of hostile action, but given the speed at which activity moves in cyberspace, these assessments may be in large part driven by computers. This raises questions as to where accountable human commanders and their staffs would be included in the important processes that support decisions to strike direct participants in hostilities.

Traditionally, actions very harmful to the interests of nations that did not involve the actual use of armed force, such as economic sanctions or espionage, were not deemed to be attacks.[95] This understanding enjoys modern currency as well, as shown by the recent definition of the crime of aggression under the Rome Statute of the International Criminal Court. "Aggression" under the Rome Statute is "the use of armed force by a State against the sovereignty, territorial integrity or political independence of another State, or in any manner inconsistent with the Charter of the [UN]."[96] Accordingly, Professor Matthew Waxman notes that "[c]omputer based espionage, intelligence collection, or perhaps even preemptive cyber operations to disable hostile systems would not constitute prohibited force, because they do not produce direct or indirect destructive consequences analogous to a military attack,"[97] that is, damage in the geophysical world.

Cyber espionage under the U.S. approach could conceivably be so significant that it would be seen as analogous to a war-like act, and under the INTERPRETIVE GUIDANCE, the required casual link could possibly be established as well. Sophisticated cyber weapons are thought to be "[c]apable of providing remarkably adaptive payloads whose activation can be triggered in milliseconds or delayed for years."[98] Further, their "[p]ayloads may even be designed to receive instructions or mutate or change their mission either by remote message or upon satisfaction of certain embedded criteria."[99] Intrusions of this sort would appear to be "significant" under the *Cyber Report*, and there could be a causal link between the espionage and the damage sufficiently direct under the ICRC position. In the end, the conclusion as to whether someone was directly participating in hostilities through conducting this sort of potential sabotage, facilitated directly by espionage, might be the same under both the INTERPRETIVE GUIDANCE and the U.S. position.

E. Perceptions of Participation

As noted *supra*, DoD is undertaking efforts to improve its ability to accurately identify actors conducting cyber operations in part through the use of "behavior-based algorithms."[100] Presumably, these algorithms would be used to evaluate how certain software had behaved and then compare these findings against criteria that reflected the identified behavioral characteristics of different actors.[101] The *Cyber Report* does not explicitly state that these algorithms can only be used to evaluate programs that had intruded into DoD systems and had been isolated –

95 Further, depending upon the circumstances, some uses of armed force between states that resulted in damage or even loss of human life have not been deemed armed conflict. Final Report, *supra* note 36, at 14, 18-19, 26-27.

96 Rome Statute of the International Criminal Court, Rome, July 17, 1998, art. 8 bis.

97 Matthew C. Waxman, *Cyber Attacks as "Force" under UN Charter Article 2(4)*, 87 INT'L LAW STUDIES 43, 48 (2011).

98 Sean Watts, *Combatant Status and Computer Network Attack,* 50 VA. J. INT'L L. 391, 402 (2010).

99 *Id*.

100 *Cyber Report, supra* note 67, at 4.

101 *See* G. Narvydas, R. Maskeliunas, & R. Raudonis, *Goal Directed, State and Behavior Based Navigation Algorithm for Smart "Robosofa" Furniture*, 10 ELEC.AND ELEC. ENG'G J. 67, 69 (2011), *available at* http://www.ee.ktu.lt/journal/2011/10/15__ISSN_1392-1215_Goal%20Directed%20State%20and%20Behavior%20based%20Navigation%20Algorithm%20for%20Smart%20Robosofa%20Furniture.pdf (schematic of algorithm in which next steps in navigation process determined in part by assessment of robot's behavior in dealing with obstacles).

perhaps they could be deployed into other digital infrastructures to examine programs resident there to determine whether they posed a threat.[102] Although it recognizes that other nations could perceive such actions as hostile, it does not appear that the U.S. believes that so doing necessarily creates a state of hostilities, or that computer operators who are conducting such intrusions are taking a direct part in hostilities.

Professor Sean Watts points out, however, that "the argument that intelligence collection, or even intelligence analysis, constitutes taking a direct part in hostilities is far stronger when such information increases the destructive effects or lethality of an attack."[103] In terms of the conduct of an actual cyber attack, if cyber specialists provide real time updates and assessments, "their contributions to the computer network attack begin […] to look progressively more like direct participation in hostilities."[104] In terms of cyber reconnaissance, the same argument holds true. The armed forces of the state whose digital infrastructure has experienced an intrusion would be derelict in their duties if they did not view that penetration as potentially destructive until shown otherwise, and even if the intrusion's initial purpose was just to find malware, it could have a secondary purpose to find a Zero Day vulnerability that could be exploited destructively at some point in the future. Arguably, the better a state conducts its cyber due diligence, the less likely it is that a mere hacker or cyber criminal could find their way into that state's digital infrastructure. Any intrusion, therefore, would likely be assigned greater seriousness simply because it occurred. This risks unnecessary and potentially unmanageable escalation.

6. CONCLUSION

The INTERPRETIVE GUIDANCE and the U.S. position represent two very different approaches to addressing the issue of direct participation in hostilities in cyberspace. The INTERPRETIVE GUIDANCE is the result of a transparent, deliberate, consensus-driven and heavily academic process geared towards ensuring the appropriate protection of civilians, consistent with its proponent's special role in promoting the continuing and enhanced observation of LOAC.[105] The U.S. position, although cognizant of the need to achieve international consensus (at least among like-minded states), is the evolving product of a nation which is at this time possibly foremost in terms of its cyber capabilities, crafted under conditions of secrecy and heavy classification while likely requiring great internal consensus among operators and civilian and military leaders, and likely geared towards preserving core U.S. economic, political and military interests. Critical examination of the two very different approaches allows an assessment of the relationships between strengths and weaknesses of each; relationships that could help define a common platform of understanding upon which to continue the discussions which must take place to determine how to apply LOAC, and in particular the concept of direct participation in hostilities, to this crucial medium of human economic, political and social interaction.

What form should these discussions take? The U.S. understandings of how it believes it would apply LOAC to operations in cyberspace may have only recently been formalized,[106] suggesting

102 Professor Melzer would argue that "probably [ ] for the purposes of targeting, data should be regarded as an object which may not be directly targeted unless it fulfills all defining elements of a military objective." Melzer, *supra* note 93, at 31.
103 Watts, *supra* note 98, at 427.
104 *Id.* at 429.
105 INTERPRETIVE GUIDANCE, *supra* note 6, at 6.
106 Waterman, *supra* note 90 (U.S. expects standing cyber ROE to be implemented by June 2012).

that there is still an opportunity for reexamination and adjustment. However, given the speed at which both cyber technology and national legal frameworks for its use appear to be evolving,[107] the ordinary process of international conferences, workshops, and meetings of experts are unlikely to prove fruitful in narrowing the gap between classified national understandings and their implementers on the one hand and public scholarly interpretations and their proponents on the other. What is needed is a common experiential approach in which national cyber security personnel, including commanders, operators and lawyers, would work together with academics and representatives of international and non-governmental organizations in cyber situational training exercises. The purpose of these scenarios would not be to test whether particular cyber strategies and tactics would be successful; rather, they would place proponents of particular legal interpretations in the position of being forced to apply those interpretations to evolving simulations. The results of the different groups working through the simulations could then be analyzed and collectively compared by the participants, and this could lead to a better appreciation on everyone's part as to how legal inputs into cyber operations might actually play out. Otherwise, the divergence between classified understandings of LOAC's application in cyber conflict and their counterparts in the public domain will likely only widen, to the detriment of defending the democratic values inherent in the notion of a cyberspace commons.

107 Ellen Nakashima, *Pentagon is accelerating development of cyberweapons*, WASHINGTONPOST.COM, Mar. 19, 2012, *available at* http://www.washingtonpost.com/world/national-security/us-accelerating-cyberweapon-research/2012/03/13/gIQAMRGVLS_story.html.

Chapter 5
“Cyber-Attacks” – Trends, Methods & Legal Classification

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Attack Trends in Present Computer Networks

Robert Koch, Björn Stelte, Mario Golling
Faculty of Computer Science
Universität der Bundeswehr München
D-85577 Neubiberg, Germany
{robert.koch, bjoern.stelte and mario.golling}@unibw.de

Abstract: An integral component of security mechanisms in company and governmental networks are Intrusion Detection Systems (IDS), which have been under intensive research for over 30 years. Unfortunately, even with these high-level security measures, the number of security incidents remains on a very high level or even rises. Therefore, for identifying the corresponding weaknesses, an in-depth knowledge of the various kinds of threats and state of the art attacks is necessary. While plenty of research about weaknesses and threats is available for special categories like wireless networks or sensor networks, research with respect to general networks, such as traditional wired networks, is widely neglected. However, the most important real-world harassment affects these networks.

In this paper we present important attack vectors based on evaluations presented in the latest technical reports, such as McAfee, M86, Symantec and corresponding academic work. For example, insider attacks and attacks on the application layer are hardly detectable by current systems, presenting challenges for intrusion detection.

To analyse the shortcomings of current IDSs, corresponding taxonomies are presented and their usability with respect to the new attack vectors is discussed. Based on this, an enhanced taxonomy is presented which addresses the current shortcomings.

Using the new taxonomy, the weaknesses of current systems are discussed, explaining the high number of serious security incidents. This knowledge can be used to design a more efficient, next-generation IDS.

Keywords: *attack trends, intrusion detection, taxonomy, next generation IDS*

1. INTRODUCTION

Current solutions for securing networks are mainly packet filters (PF), application layer gateways (ALG) and IDS. PF and ALG are used to control traffic that enters a network and leaves a network based on packet information. They filter malicious network traffic according to predefined rule sets. Known shortcomings of PF and ALG are generally [1,2]:

- They cannot protect against attacks that bypass them, such as tunnelled traffic.
- They do not protect against threats caused by internal attackers.
- They hardly protect against the transfer of malicious code.

To overcome some of these shortcomings IDSs are used in combination with PFs and ALGs. IDSs are primarily for learning, detecting and reporting attacks as they happen in real time. Basically, two types of IDS are available: signature-based (misuse) and statistical-based (anomaly) detection. Signature-based IDSs use pattern matching to detect signature traces in network traffic. A detection of attacks is only possible for known attack signatures. Signature-based IDSs are considered to have a low false positive, but unfortunately a relatively high true negative, detection rate. In contrast, anomaly-based IDSs are able to detect new kinds of attacks but at the price of higher false positive rates. State-of-the-art IDSs are based on traditional taxonomies which hardly reflect recent attack vectors. Based on recent reports, such as [3-7], we have identified important threats for security solutions of traditional networks.

In Section 2 we will present these threats, which are application layer attacks, zero-day exploits, social engineering, targeted attacks, dissemination routes, data leakage and insider attacks, encryption, IPv6 attacks and attacks on and with the use of cloud computing. A taxonomy for intrusion detection is presented in Section 3 and the shortcomings of current systems are discussed. Finally, the paper is concluded in Section 4.

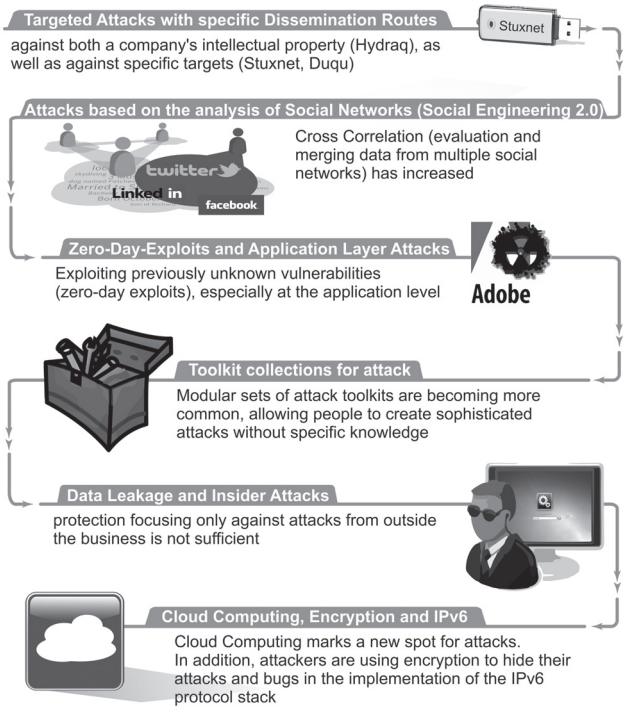
2. ATTACK AND TECHNICAL TRENDS

Nowadays, economic crime affects many large companies [8]. 61 percent of these companies reported that they have become the victims of economic crimes in the past two years. On average, these companies report eight cases a year. In addition to high financial losses, substantial non-pecuniary damage is reported: loss of reputation, damage to business and loss of morale.

Symantec recorded more vulnerabilities in 2010 than in any previous year since starting their internet security thread report[6]. While many attacks are directed at large enterprises and governmental organisations, they can also target small and medium businesses and individuals. Similarly, senior executives are not the only employees being targeted. In most cases, a successful compromise requires only victimising a user with just limited access to network or administrative resources. A single negligent user or unpatched computer is sufficient to give attackers a beachhead into an organisation from which to mount additional attacks on the enterprise from within, often using the credentials of the compromised user [34].

Based on annual security reports from Panda [4], McAfee [9], M86 Security [10] and Symantec [5-7], we have identified the following attack trends, summarised in Figure 1 and discussed in the next paragraphs: application layer attacks, zero-day exploits, social engineering, targeted attacks, dissemination routes, data leakage and insider attacks, encryption, IPv6 attacks, and attacks on and with the use of cloud computing.

FIGURE 1. TODAY'S ATTACK TRENDS



A. Application Layer Attacks

Each communication layer has its own security challenges. In particular, the application layer with its variety of supported protocols offers many vulnerabilities and access points for attackers and in return makes it very difficult to fend off attacks. Furthermore, attacks on this layer are especially attractive to attackers, since this layer offers direct access to information without for example the need for a cumbersome extraction of the payload from the package.

Botnets are one of the most important security harassments today. Numerous systems like personal computers are misused and remote-controlled by the installation of local agents. Because of their placement within the network (which is typically secured against access from outside), Bots are able to communicate to an external server taking commands and executing attacks. Botnets are hard to detect for traditional IDSs, but even more complex because of encrypted communication methods and distributed control systems of modern botnets. Another important fact is that more than 70% of the current attacks are conducted on the application layer. Therefore, they have to be evaluated using the packet payload. On the other side, encryption technologies like TLS are more and more widespread, hampering the application of payload inspection methods (deep packet inspection, DPI).

Due to increasingly complex application software, like browsers with their numerous Add-Ons, extensive vulnerabilities are available and used intensely by attackers.

Some of the relevant application layer attacks are [7]:

- Scripting vulnerabilities
- Cookie poisoning
- Hidden field manipulation
- Parameter tampering
- Cross-site scripting
- SQL injection

In addition, traditional attack techniques like buffer overflows are also used to execute attacks on the application layer. Even techniques like address space layout randomization (ASLR), which makes sure that system functions are located at randomly chosen addresses (instead of being located at the same memory address anytime), or sandboxing can be overcome by sophisticated attacks like JIT-spraying. According to security reports published by Symantec, since a few years ago the proportion of application layer attacks is over 70% in comparison with the total amount of all attacks, and still increasing, therefore displacing traditional operating system and network layer-oriented attacks [6,7,11].

B. Zero-Day Exploits

A Zero-Day exploit occurs when a flaw is discovered in software and a programme exploiting the vulnerability is available before or on the day the vendor gets to know about the flaw.

So-called "file format vulnerabilities" remain the first choice for Zero-Day exploits. In this regard, most attacks relate to Adobe PDF, Flash Player and Microsoft Office Suite (PowerPoint, Excel and Word) and the corresponding third party add-ons (which make the patching process more complicated and thus increases the options for potential attackers). The time vendors need for developing patches against Zero-Day vulnerabilities is often too long, for instance because they want to stick to so-called Patch-Days instead of releasing updates individually. Often vendors are unable to fix vulnerabilities quickly due to a lack of security-by-design. In software engineering, this means that the software has to be designed from the ground up to be secure.

All in all, Zero-Day vulnerabilities remain one of the major threats and, therefore, require additional security measures.

C. Social Engineering

Social engineering was used intensely in the 1980s, for example by well-known hackers like Kevin Mitnick. Social engineering describes a non-technical kind of intrusion that relies heavily on human interaction and often includes fooling other people to dig normal security measures. Social engineering is a key component for today's and upcoming attacks, utilising the weakest link of the chain, the user. In distinction to other technical measures, here, attackers may seem unassuming and respectable (possibly a new employee or repairman and sometimes even with some credentials to prove the faked identity).

One of the most known attacks in the field of social engineering is called the phishing attack.

It uses emails or malicious websites to gather personal information by claiming to be a trusted organisation.

Other techniques, such as scareware, rogueware, and ransomware-attacks, are also known. Scareware includes several types of scam software with malicious payloads, or limited or no benefit, often sold to consumers by unethical business practices. The approach uses social engineering to cause shock, fear or the perception of a threat, usually to unsuspecting users. Rogueware is a form of computer malware that causes users to pay money for the faked or simulated removal of malware. Ransomware is computer malware that holds a computer system, or the data contained therein, as a hostage to its users with a demand for ransom for the restoration.

Awareness of the risks and available safeguards is the first line of defence for security of information systems and networks. Some problems which need to be addressed in the field of social engineering are [12]:

- People do not understand the technology
- People are caught off guard
- People trust known people (co-workers)
- People trust the system
- People are in a hurry
- People get careless

D. Targeted Attacks

The times of large-scale virus attacks have mostly passed. Some of the biggest threats to the security of corporate networks nowadays are targeted attacks. Here, in contrast to other attacks, the design is specifically tailored to individuals or organisations. Thus, on the one hand the probability that the victim actually opens the e-mail is increased and, on the other hand, existing protective measures are easier to be bypassed. Therefore, the attacker starts with identifying potential victims by making use of public available data like the website of a company or the data available in social networks like Facebook or Twitter. Many people are careless when dealing with sensitive data, especially in the context of social networks. Due to the personal data found in the network an individualised email concerning a current topic and containing a malicious payload is generated and sent to the victim. If the victim opens the payload, than the computer can be used and controlled by the attacker.

Since 2005, an increase in targeted attacks on federal agencies and industrial espionage can be observed [8]. Public attention was especially gained in 2007, when numerous computers in federal ministries and the German Chancellery were infected with spy-ware as a result of a targeted attack. Recently, some methods have emerged that allow an even more sophisticated profiling, enabling an attacker to start more advanced targeted attacks or to improve the efficiency of spam campaigns. Here, the profiles of the different social networks are evaluated by special procedures and automatically linked between each other to enrich the information (cross-correlation). It has been demonstrated that – based on a list of about 10.4 million e-mails – the automatic user profiling of more than 1.2 million user profiles, including the linking between different social networks, is possible.

Other important examples of targeted attacks are the Hydraq Trojan (also known as Aurora) which affected Google and several other large companies in 2009, or the attack on RSA in 2011, which was compromised by attackers using this Trojan [13]. In Aurora, a Zero-Day vulnerability which affected three versions of Internet Explorer and various Windows operating systems was used. The attackers sent targeted emails to people of high-ranking management who had privileged access rights to various applications [4]. Afterwards, the malicious code was used to access and steal information from Gmail accounts. The attack on RSA and the consecutive attacks on Lockheed Martin and other US defence contractors are some of the latest and most sophisticated examples of a targeted attack. First, the network of RSA was attacked by the use of social networks and a vulnerability in Adobe Flash [14]. In the next step, data about employees of the company was collected and used to send personalised phishing emails. The emails contained a malicious spreadsheet which exploited a Zero-Day vulnerability in Adobe Flash and enabled remote access to the attackers. By that, information about 40 million two-factor authentication accounts of SecureID was stolen. After that, malware and phishing attacks were used to link tokens to end-users [3]. Based on this association, the consecutive attacks on Lockheed Martin and other companies were carried out by compromising the SecurID accounts.

E. Dissemination Routes

The dissemination routes of malicious software are not restricted to networks like the Internet or services like email. Just like at the beginning of the development of malicious software in the mid-80s, data storage media is an important method of distribution. The formerly used floppy disks have been replaced by cheap memory sticks with high capacity. Because of the use of the autorun-functionality, an infection can be automated easily. For example, promotional gifts like USB-sticks given away at trade shows are popular instruments [34]. By connecting the stick to a computer, a Trojan – previously placed on the stick – installs itself onto the system [34]. Therefore, malicious code is injected directly into the target system or inside a network, bypassing the security systems.

With the help of this offline-propagation method, formerly secure systems and networks like Supervisory Control And Data Acquisition Systems (SCADA) can also be compromised, as demonstrated by the well-known example of Stuxnet [34].

In addition, the attack tool's automation level and their sophistication continue to improve. No technical in-depth knowledge is needed to create new, unknown and malicious software any longer [15]. The first attack kit named Virus Creation Lab in 1992 provided basic functionality, but state-of-the-art kits like Mpack and Nukesploit or Command-and-Control toolkits such as Spy Eye or Zeus are highly professional [16]. These toolkits are sold for several thousand Dollars with different service levels. Due to their professionalisation and commercialisation, these easy-to-use attack kits can produce serious damage.

F. Data Leakage and Insider Attacks

The term data leakage prevention (DLP) refers to the protection against a suspected, but not measurable and sometimes not even detectable, sharing of information to unwanted recipients [17]. In contrast to insider threats, data leakage includes accidental or unintentional data loss in addition to malicious theft [18]. Numerous scandals about data loss and data theft have gained public interest in the recent past [19, 20]. While governments and militaries were in

the spotlight of attacks during the cold war, today, the industry is the most important target for espionage. For example, a study of the consultancy PricewaterhouseCoopers and the University Halle-Wittenberg specified that the economic loss for each individual business company in Germany was on an average about 5.57 million Euros in 2009. Sixty-one percent of all large-scale enterprises had been hit by business crime in the past two years [8].

Regarding the protection against industrial espionage and information flows out of the company, many businesses focus only on protection against attacks from outside. In times of rising fears of losing one's job, permanently growing workloads and often a lack of appreciation of performance, many employees are increasingly willing to enrich themselves at the expense of the company they are working for. Loyalty to the employer is no longer always natural. A loss of wages is thus more often compensated by a small additional income. The particular endangerment by the insider is based on the authorised access and their knowledge about the security mechanisms. The numbers of insider threats compared with all incidents of data loss differ keenly from 17% up to 80% [21, 22]. When investigated in a study conducted of German companies about types of employees who were specifically responsible for the espionage, first and foremost, the clerks (who usually have many access rights including access to sensitive documents and information) with 31.4% were identified, followed by skilled workers with 22.9% and 17.1% within the management. Together these three areas caused about two thirds of the entire data leakage of the company.

Countermeasures to avoid data leakage are quite complex. For example, all files that are read by or written to all USB devices must be logged so that each change to sensitive data is traceable. Furthermore, with the use of a unique serial number, a USB stick can be assigned to only one specific user. As the stick is encrypted, reading the data on the stick is only possible for colleagues of the department or superiors.

G. Encryption

Cryptography was invented to protect communication; this is the reason why militaries in the world and scientists have developed it. Even the protection of stored data can be seen as a form of communication [23], here with the addition that each key must exist as long as the encrypted data exists. The storage of these keys is thus as important as the storage of the data. Therefore, encryption is not reducing the number of secrets that must be stored safely; it is only making their size smaller. In the past, keys have been stored in the human brain and by that in a way that is not connected to a network (and thus kept safe in principle), but this approach does not work for the Internet today. Often, keys are needed for the communication between systems in an automated fashion; shared secrets must be stored, etc. So, keys can no longer be saved in the brains of people. They must be stored on the same computer that hosts the data or at least on a network-wide available system – and that is a lot riskier.

Beside the challenges regarding the security of the keys itself, the usage of encryption rises generally. Not only are more and more services and servers offering encrypted access to their customers, the attackers are also making increasing use of cryptography to hide and to secure their activities. For example, the latest botnets use encrypted communication channels to hide their presence from IDSs or next generation firewalls.

The trend towards the use of encryption will also be enforced with the broader application of IPv6. This will have a significant impact on the applicability of security devices and mechanisms and the detectability of attacks.

H. IPv6

At the moment, due to missing IPv6 security features in routers, firewalls and other critical network infrastructures and the lack of IPv6 testing and experience, many Internet providers tend to slowly migrate from IPv4 to IPv6, or at least they deploy IPv6 parallel to IPv4. A recent study showed a percentage of IPv6 traffic of just 0.03% compared with 0.002% from the previous year [24]. Nevertheless, the amount of IPv4-to-IPv6 tunnels will increase and it is still not clear whether all of these tunnels are implemented correctly. Some have the view that attacks that make use of IPv4-to-IPv6 tunnels to conceal the attack are already known.

IPSec is a mandatory component of IPv6 and is implemented using the authentication header and the Encapsulating Security Payload extension header [25]. In February 2011, the last address block of IPv4 was assigned [26]. The lack of IPv4 addresses on the one side and the increasing number of new devices on the Internet on the other side, for example mobile devices like smartphones, will speed-up the utilisation of IPv6 in the near future and, therefore, also the even wider distribution of encryption as mentioned above.

I. Cloud Computing

Many people and organisations are nowadays using cloud services to take advantage of convenience and attractive pricing (e.g., pay-as-you-go financing).

Nevertheless, there are valid security concerns including lack of control of data, downtime due to an outage and lack of visibility as already outlined in [33]. Despite excellent security practices employed by many cloud providers, the fact remains that these services are likely to be prime targets. During 2011, as mentioned in [10], Sony's PlayStation network was hacked, leading to a shutdown in the service that affected about 77 million users. LastPass, a web-based password management company, also had its system breached, resulting in the necessity for all users to change master passwords [33]. Cloud service providers are huge targets. Since the data is concentrated, and the systems are standardised, a successful breach could yield a lot of valuable data for an attacker. For these reasons, it is predicted that more high-profile attacks on cloud service providers are to come in 2012 [10].

In addition it has been demonstrated, for instance in [35], that attackers can make use of cloud services for purposes like breaking encryption using tools, like the so called 'Cloud Cracking Suite'.

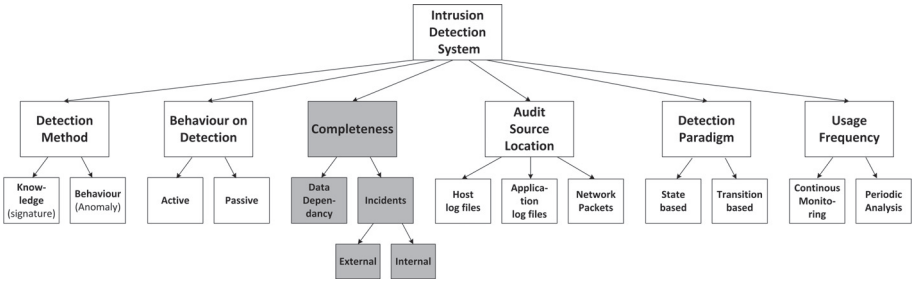
3. TAXONOMY

To understand the origin of the security incidents respectively the shortcomings in the detection process, current IDSs, can be evaluated by a taxonomy. A classification or taxonomy is a hierarchical structure of a field of knowledge into groups. The classification has to be made in a manner that several properties are satisfied (e.g., [27, 28]): mutual exclusiveness, completeness,

traceability, convenience, clarity and acceptance. The divisions are used, for example, to investigate new systems ordered by weaknesses.

No generally accepted taxonomy is available for the allocation of IDSs, therefore, various classifications of very different levels of detail can be found in the literature. The taxonomy published by Debar et al. [29] is widely used: The most common characteristic is the detection method, behaviour- (anomaly) or knowledge- (signature) based. The behaviour of detection can be active or passive while the audit source location can be the host or application log files or network packets. While the detection paradigm can be state- or transition-based, the frequency of usage can be continuous monitoring or a periodic analysis. With respect to present attacks and technical trends, this division is not sufficient for the analysis and evaluation of today's IDSs. Especially, the aspect of the completeness of the evaluation must be taken into consideration for modern systems because of the increasing amount of traffic which is encrypted and is, therefore, not analysable by most IDSs. Because of this, we enhanced the current taxonomies with the category "completeness" and the corresponding sub-levels "data dependency" and "incidents" (see Figure 2). While the first one describes the dependency of an IDS to have access to the communication data, especially the payload of the network packets, the latter one considers the detectability of attack sources, both external and insiders.

FIGURE 2. CLASSIFICATION OF IDS BASED ON THE TAXONOMY OF DEBAR AND EXTENDED WITH THE CATEGORY COMPLETENESS



Also, the categories like social engineering, targeted attacks or insider attacks are not represented. Even though this is not the original goal of the (technical) IDS taxonomies, these are the most important attacks today, thus having an high influence on the assessment of IDSs. Numerous other taxonomies exist, e.g. the comprehensive classification by Sabahi and Movaghar [30], which includes aspects like the environment, or the taxonomies of Sundaram [31] and Bolzoni [32] which are specialised on a fine subdivision of the detection behaviour respectively, a specialised taxonomy for anomaly-based systems. Anyway, none of the existing taxonomies are able to reflect the current attack trends and, therefore, cannot give a meaningful evaluation of the performance of today's IDSs.

Regarding the attack and technical trends identified, the following requirements must be fulfilled by a modern IDS:

- A complete analysis of the network traffic must be provided, independent from the data layer and from protective measures like payload encryption.

- Several characteristics of modern attacks exploit human weaknesses, e.g. when using social engineering methods or targeted attacks. Also bots, which are installed on systems of e.g. a company network (therefore, already inside the trusted network) or activities by insiders can be difficult to detect. These are properties which can be hardly detectable on a technical layer. Therefore, the capabilities of an IDS must comprise detection methods for attacks from the outside as well as irregularities of any kind from the inside, which typically will only possible by sophisticated anomaly-based techniques.

The shortcomings of current taxonomies and the abstract enhancement with the category "completeness" emphasise the central challenge of today's intrusion detection: the sophisticated and advanced attack techniques make use of all levels of abstraction – from technical aspects to human weaknesses. Therefore, some important attack vectors are hardly detectable with purely technical procedures. Behaviour-based detection systems are mandatory to overcome the current shortcomings, but also with these techniques, the completeness of the detection with respect to the possible attacks remains a crucial factor which has to be evaluated for every IDS in depth.

4. REMARKS

As already reported in [33], the Internet has revolutionised our social and business habits today. It has evolved from a network of computers and information into a network of people. The future Internet will consist of dynamically scalable and virtualised resources, which will be provided by providers as a service over the Internet. Aside from the obvious socio-economic aspects, also the technical side will change considerably. Due to the fact that the number of "services over the Internet" will increase tremendously and get more and more important as new business models, the providers of the future Internet will need to cope with new problems.

They will not only have to solve scalability and availability problems, but more importantly new security issues will arise and so new kind of attacks on the future Internet will be feasible. Key challenges in such a highly complex environment where data and services are also located somewhere in "clouds" are security, privacy and trust. The term "services over the Internet" implies that not only the data of the end users has to be encrypted, but also the whole network communication from end user to service providers. This claim for encryption is not only to justify the end user acceptance of services. Legally, regulations like BASEL II and most EU and national data privacy laws mandate that firms are to encrypt information transferred over the network when using services provided in the future Internet [33].

The emergence of new technologies and services, as well as trillions of devices and petabytes of data to be processed and transferred, mean that we have to deal with new threats and vulnerabilities, in addition to handling the remaining old ones. One must cope with attacks on the networks, but well-established IDSs and Internet early warning systems (IEWS) will not defend anymore, because of the encrypted packet payload [33]. The provider has no chance to decrypt the packet payload since the decryption key is not available and de- and encryption of millions of packets is too resource devouring.

Since neither the Internet, nor the future Internet consist only of national networks and national providers, the described problem needs to be addressed on a multinational level. Services are already offered nowadays to the end users without the information where the services or parts of the services are located (e.g. cloud computing). Nor is the end user interested in the service location but only in the availability and the safeguarding of the service. National and international providers need expert knowledge in how to secure provided services to end users and how to detect and prevent next-generation networks attacks.

5. CONCLUSION

Even if firewalls and state-of-the-art IDSs are in place in today's company networks, the number of incidents remains on a high level and new incidents are reported on a day-to-day basis. Several aspects have been identified which are responsible for the bad performance of current security systems: more and more attacks are targeted attacks and specifically designed and social engineering is used to bring the victim to execute the malicious operation. By the use of, for example, memory sticks, secured and isolated systems and networks can also be attacked. Application layer attacks, an increasing number of Zero-Days and the insider threat are further tendencies. The specific characteristics of these trends cannot be reflected by current taxonomies, therefore, hampering the development of new security systems and devices. The human being remains the weakest link of the chain, enabling sophisticated attacks where the legal user is manipulated to execute the disguised attack by himself with his authorised access and without realising the subjacent attack. To overcome these shortcomings, new concepts for the support and comprehension of users into the security processes are necessary.

REFERENCES:

[1] S. Jin, Y. Wang, X. Cui, and X. Yun, "A Review of Classification Methods for Network Vulnerability," In *Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, October 2009, pp. 1171–1175.

[2] V. M. Igure and R. D. Williams, "Taxonomies of Attacks and Vulnerabilities in Computer Systems," In *IEEE Communication Surveys, ser. IEEE Communication Surveys and Tutorials*, IEEE, 2008, vol. 10, pp. 6–19.

[3] Reuters, *Hackers breached us defense contractors*, [Online]. Available: http://www.terminalx.org/2011/05/hackers-breached-us-defense-contractors.html#axzz1QCJWtZJz, May 2011.

[4] Pandasecurity, *Annual report PandaLabs 2010*, Tech. Rep., 2010. [Online]. Available: http://www.pandasecurity.com.

[5] M. Fossi et al., *Symantec Global Internet Security Threat Report*, Symantec Corporation, Tech. Rep. XV, April 2010.

[6] M. Fossi et al., *Symantec internet security report trends for 2010*, Symantec Corporation,350 Ellis Street, Mountain View, CA 94043 USA, Tech. Rep., April 2011.

[7] M. Fossi et al., *Symantec internet security report trends for 2011*" Symantec Corporation, 350 Ellis Street, Mountain View, CA 94043 USA, Tech. Rep., April 2012.

[8] K.-D. e. a. Bussmann, *Wirtschaftskriminalität 2009*, Pricewaterhouse-Coopers, Martin-Luther-Universität Halle-Wittenberg, Tech. Rep., September 2009.

[9] M. Labs, *McAfee threat predictions 2012*, McAfee Corporation, 2821 Mission College Boulevard, Santa Clara, CA 95054 USA, Tech. Rep., Jan 2012.

[10] M86 Security, M86 *Security Labs: Thread predictions 2012*, 8845 Irvine Center Drive, Irvine, CA 92618 USA, Tech. Rep., Jan 2012.

[11] M. Fossi, *Symantec internet security report trends for 2009*, Symantec Corporation, 350 Ellis Street, Mountain View, CA 94043 USA, Tech. Rep., April 2010.
[12] ENISA, *A Users' Guide: How to Raise Information Security Awareness*, European Network and Information Security Agency, Tech. Rep., 2006.
[13] EMC, *Open Letter to RSA SecurID Customers*, [Online] Available: http://www.rsa.com/node.aspx?id=3891.
[14] M. Kobie, *Rsa blames flash flaw and social media for attack*, [Online] Available: http://www.pcpro.co.uk/news/security/366532/rsa-blames-flash-flaw-and-social-media-for-attack.
[15] J. McHugh, A. Christie, and J. Allen, *"Defending yourself: The role of intrusion detection systems"*, in *Software*, IEEE, 2000, vol. 17, no. 5.
[16] M. Fossi, *Symantec Report on Attack Kits and Malicious Websites*, Symantec Corporation, Tech. Rep., 2010.
[17] P. Papadimitriou and H. Garcia-Molina, "Data leakage detection," in *Knowledge and Data Engineering, IEEE Transactions on*, vol. 23, no. 1, pp. 51 –63, Jan. 2011.
[18] M. McCormick, "Data Theft: A Prototypical Insider Threat," in *Advances in Information Security*, vol. 39, no. 1, pp. 53–68, April 2008, ISBN-10:0-387-77321-5.
[19] M. Simons, *Ministry of Defence in new data loss scandal*, [Online] Available: http://www.cio.co.uk/news/3225/ministry-of-defence-in-new-data-loss-scandal, October 2008.
[20] Backup-Technology, *Data loss incident affects NASA*, [Online] Available: http://www.backup-technology.com/5451/data-loss-incident-affects-nasa/, December 2010.
[21] KPMG, *e-Crime-Studie 2010*, [Online]. Available: http://www.kpmg.de/Themen/21481.htm, 2010.
[22] W. e. a. Baker, *2011 Data Breach Investigations Report*, Verizon Business, Tech. Rep., [Online] Available: http://www.verizonbusiness.com, 2010.
[23] B. Schneier, *Secrets and lies: digital security in a networked world*, John Wiley, 2000.
[24] C. Labovitz, *Six Months, Six Providers and IPv6*, [Online] Available: http://asert.arbornetworks.com/2011/04/six-months-six-providers-and-ipv6/, 2011.
[25] D. Kaushik, *Ipsec & ipv6 - securing the nextgen internet*, [Online] Available: http://ipv6.com/articles/security/IPsec.htm, 2008.
[26] M. Ermert, *Ipv4-adressen: Abschiedsgrüße, Mahnungen und Pappschilder*, [Online] Available: http://www.heise.de/netze/meldung/IPv4-Adressen-Abschiedsgruesse-Mahnungen-und-Pappschilder-1183204.html, 2011.
[27] J.D. Howard and T.A. Longstaff, *A Common Language for Computer Security Incidents*, Technical report, Sandia National Laboratories, Albuquerque, New Mexico 87185 and Livermore, California 94550, 1998.
[28] S. Jin, Y. Wang, X. Cui, and X. Yun, "A Review of Classification Methods for Network Vulnerability," in *Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics*, pages 1171–1175. IEEE, Oktober 2009.
[29] H. Debar, M. Dacier, and A. Wespi, *"A Revised Taxonomy for Intrusion-Detection Systems"*, Technical report, IBM Research, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland, 1999.
[30] F. Sabahi and A. Movaghar, "Intrusion Detection: A Survey," in *Systems and Networks Communications*, 2008. ICSNC '08. 3rd International Conference on, pages 23–26. IEEE Computer Society, Oktober 2008. DOI 10.1109/ICSNC.2008.44.
[31] A. Sundaram, *An introduction to intrusion detection*, Crossroads, 2(4):3–7, April 1996. DOI: http://doi.acm.org/10.1145/332159.332161.
[32] D. Bolzoni. *Revisiting Anomaly-based Network Intrusion Detection Systems*. PhD thesis, University of Twente, 2009. DOI: 10.3990/1.9789036528535.
[33] M. Golling and B. Stelte, "Requirements for a future EWS-Cyber Defence in the internet of the future," In *3rd International Conference on Cyber Conflict (ICCC)*, IEEE, 7-10 June 2011, pp. 135–150.
[34] R. Koch, "Towards next-generation Intrusion Detection," In *3rd International Conference on Cyber Conflict (ICCC)*, IEEE, 7-10 June 2011, pp. 151-168.
[35] T. Roth, *Breaking encryption in the cloud: GPU accelerated supercomputing for everyone*, Black Hat DC 2011, [Online], Available: http://blackhat.com/html/bh-dc-11/bh-dc-11-briefings.html.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

"Attack" as a Term of Art in International Law: The Cyber Operations Context

Michael N. Schmitt
International Law Department
United States Naval War College
Newport, U.S.A.
schmitt@aya.yale.edu

Abstract: This article examines the meanings of "attack" in international law. It points out that the term is used in two distinct bodies of that law. First, the term "armed attack" appears in the *jus ad bellum*, which governs when a State may resort to force as an instrument of its national policy. In that context, it serves as a condition precedent to the resort to force in self-defence pursuant to Article 51 of the UN Charter and customary international law. Second, in the *jus in bello* attack refers to a particular type of military operation to which various prohibitions and restrictions apply. The *jus in bello*, or international humanitarian law, establishes rules as to how operations may be conducted during an armed conflict. The article examines and analyses these usages both to distinguish them from each other and to better inform the non-legal community as to their legal significance.

Keywords: *jus ad bellum, jus in bello, international humanitarian law, armed attack, self-defence, attack, distinction*

1. INTRODUCTION

The U.S. Department of Defense's *Dictionary of Military Terms* defines "computer network attack" (CNA) as "[a]ctions taken through the use of computer networks to disrupt, deny, degrade, or destroy information resident in computers and computer networks, or the computers and networks themselves."[1] NATO adopts this definition in its *Glossary of Terms*, but adds the parenthetical that "[a] computer network attack is a type of cyber attack."[2] Curiously, it does not define "cyber attack" and the reference contains the sole mention of "cyber" in the document.

The term "computer network attack" is adequately descriptive for non-legal use. For instance, it usefully distinguishes such operations from computer network defence, computer network

1 U.S. Department of Defense, Dictionary of Military and Associated Terms, Joint Publication 1-02, Nov. 8, 2010, as amended through Feb. 15, 2012, available at http://www.dtic.mil/doctrine/dod_dictionary/.
2 NATO Standardization Agency, NATO Glossary of Terms and Definitions (AAP-6) (2010), at 2-C-12.

exploitation and other cyber activities.[3] Despite practical utility, its use causes measurable disquiet among lawyers, for "attack" is a legal term of art that has specific meaning in the context of two very different bodies of international law governing State behaviour in times of crisis or conflict. In both cases, the term represents a consequential threshold that delineates the legality of particular cyber operations, and, in some cases, the lawfulness of responses thereto.

This article seeks to bridge the terminological gap between the legal and non-legal communities by examining and explaining the significance of the word "attack" in international law. Hopefully, doing so will imbue policy makers, cyber operators and technical experts with greater sensitivity to the legal dimensions of the verbiage they employ when addressing cyber matters. Although the two communities may not speak the same language, members of both benefit from being bilingual.

2. THE LEGAL ARCHITECTURE

The international law governing conflict consists of two distinct bodies of law: the *jus ad bellum* and the *jus in bello*. *Jus ad bellum* norms govern when States, as an instrument of their national policy, may resort to force. They address, inter alia, the prohibition of the use of force by States and the exceptions thereto, most notably the right of self-defence and authorization or mandate by the UN Security Council.[4] The *jus in bello*, by contrast, deals with how the military and other armed actors may employ force, including who and what may be targeted.

These norms, also labelled the "law of armed conflict" or "international humanitarian law" (the latter term adopted in this article), apply in situations of "armed conflict" irrespective of whether the State or armed actor in question has resorted to force in compliance with the *jus ad bellum*. Differing objects and purposes animate the two bodies of law and explain the impenetrable barrier between them. The *jus ad bellum* seeks to maintain peaceful relations within the community of nations by setting strict criteria as to when States may move beyond non-forceful measures such as diplomacy, economic sanctions and counter-measures.[5] Of particular note is the right to do so in self-defence when either facing an "armed *attack*" or coming to the aid of another State which is defending itself (collective self-defence). By

3 Computer network operations comprise "computer network attack, computer network defense, and related computer network exploitation enabling operations. DoD Dictionary of Military Terms, *supra* note 1. Computer network defense is defined as "[a]ctions taken to protect, monitor, analyze, detect and respond to unauthorized activity within Department of Defense information systems and computer networks," whereas computer network exploitation encompasses "[e]nabling operations and intelligence collection capabilities conducted through the use of computer networks to gather data from target or adversary automated information systems or networks." *Id*.

4 U.N. Charter, arts. 2(4), 42 & 51.

5 Countermeasures are "measures that would otherwise be contrary to the international obligations of an injured State vis-à-vis the responsible State, if they were not taken by the former in response to an internationally wrongful act by the latter in order to procure cessation and reparation." Draft Articles on Responsibility of States for Internationally Wrongful Acts, Report of the International Law Commission on the Work of its 53rd sess., UNGAOR, 56th sess., sup. No. 10 (A/56/10), ch. IV.E.1, at p. 128, *available at* http://untreaty.un.org/ilc/texts/instruments/english/commentaries/9_6_2001.pdf [hereinafter Articles of State Responsibility]. Note that Article 50 of the Articles of State Responsibility provides that countermeasures cannot amount to a use of force. However, this position, which the author accepts, was challenged by Judge Simma in the Oil Platforms case, where he argued that countermeasures could involve force when in response to an act that itself amounted to a use of force, but did not qualify as an armed attack. Oil Platforms (Islamic Republic of Iran v. U.S.), 2003 I.C.J. 161, ¶¶12-13 (Nov. 6) (separate opinion of Judge Simma).

contrast, international humanitarian law seeks to minimize harm during an armed conflict that is either unnecessary to effectively accomplish legitimate military aims or excessive relative to them. It does so most directly by establishing legal boundaries for the conduct of "*attacks*." Ignoring "right or wrong" under the *jus ad bellum* optimizes this purpose.

Since the term "attack" applies in separate bodies of law with discrete objects and purposes, it is unsurprising that its meaning differs depending on its source. In the *jus ad bellum*, it appears in Article 51 of the United Nations Charter: "Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security." Article 51, recognized as reflective of customary international law by the vast majority of legal scholars, is an express exception to Article 2(4) of the Charter, which provides that "[a]ll Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations." Taking the Articles together, a State may "use force" without violating Article 2(4) when it is the victim of an "armed attack", as that term is envisaged in Article 51. Self-defence requires no *ex ante* authorization from the Security Council, States alone enjoy the right of self-defence, and the right only attaches to armed attacks with a transnational element.[6]

In international humanitarian law, "attack" refers to a particular category of military operations. Article 49(1) of the 1977 Additional Protocol I to the 1949 Geneva Conventions defines "attacks" as "acts of violence against the adversary, whether in offence or in defence."[7] It is a neutral term in the sense that some attacks are lawful, whereas others are not, either because of the status of the object of the attack or how the attack is conducted. Neutral though it may be, "attack" is operatively a key threshold concept in international humanitarian law because many of its core prohibitions and restrictions apply only to acts qualifying as such.

It is important to bear in mind that this notion only attains relevance once an "armed conflict" is underway. Like "attack", "armed conflict" is a legal term of art referring to two types of conflicts: 1) international armed conflicts, which are between States; and 2) non-international armed conflicts, which are conflicts at a certain level of intensity and organization between a State and an organized armed group or between organized armed groups.[8] Absent a situation qualifying as one of these conflicts, domestic and human rights law, not humanitarian law, governs the activities in question.

6 In the cyber context, the meaning of the term "use of force" is highly unsettled. See Manual on the International Law Applicable to Cyber Warfare (Tallinn Manual), (Michael N. Schmitt et al. eds., Cambridge University Press, forthcoming 2013) [hereinafter Tallinn Manual].

7 Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, art. 49.1, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter AP I].

8 For the thresholds applicable to international and non-international armed conflict, see common articles 2 and 3 respectively to the four Geneva Conventions. Note that in addition to situations involving hostilities, the applicability of humanitarian law extends to those in which there has been a declaration of war or occupation, even when hostilities have not broken out. Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, Aug. 12, 1949, 6 U.S.T. 3114, 75 U.N.T.S. 31; Convention for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea, Aug. 12, 1949, 6 U.S.T. 3217, 75 U.N.T.S. 85; Convention Relative to the Treatment of Prisoners of War, Aug. 12, 1949, 6 U.S.T. 3316, 75 U.N.T.S. 135; Convention Relative to the Protection of Civilian Persons in Time of War, Aug. 12, 1949, 6 U.S.T. 3516, 75 U.N.T.S. 287.

To summarize, an "armed attack" is an action that gives States the right to a response rising to the level of a "use of force," as that term is understood in the *jus ad bellum*. By contrast, the term "attack" refers to a particular type of military operation during an armed conflict to which particular international humanitarian law norms apply. The general outline fashioned, it is apropos to examine the terms as they apply in the cyber environment.

3. CYBER "ARMED ATTACKS" UNDER THE JUS AD BELLUM

Before turning to the possible qualification of cyber operations as armed attacks, it is important to grasp the related point that there are no unique restrictions on the resort to defensive cyber operations in response to kinetic operations that qualify as an armed attack. On the contrary, they mirror those applying to kinetic defensive actions. For instance, cyber operations have to comply with the *jus ad bellum* principle of necessity, by which force may only be employed defensively to the extent non-forceful measures are unlikely to suffice. They equally have to comport with the *jus ad bellum* principle of proportionality, allowing only that degree of force required for an effective defence.[9] Cyber uses of force in the face of an armed attack must further meet the related requirements of imminency and immediacy, which limit, respectively, responses in anticipation of, and subsequent to, an attack. These and other questions, in particular the legal meaning of the phrase "use of force", are dealt with at length in the forthcoming *Tallinn Manual*.[10]

The question at hand, however, is when does a cyber operation qualify as an armed attack, that is, when does an action against a State legally merit a response with either cyber or kinetic actions that are at the level of a use of force?[11] The challenge lies in interpreting the adjective "armed." "Armed" is not to be equated with "force" in the sense of Article 2(4). The International Court of Justice recognized this normative "gap" in the *Nicaragua* Judgement when it found that there are "measures which do not constitute an armed attack but may nevertheless involve a use of force" and distinguished "the most grave forms of the use of force from other less grave forms."[12] The Court cited supplying weapons and providing logistical support to a rebel group in another State as an example of a use of force that did not amount to an armed attack against that State.[13] This gap makes sense in light of the central object and purpose of the United Nations Charter – to craft a system that effectuates a strong presumption against the use of force in international relations and favours collective responses to threats to (or breaches of) the peace over unilateral ones.

The result is a normative schema in which all armed attacks are uses of force, but not all uses of force are armed attacks. As a consequence, States may face cyber operations constituting a use of force, but be unable to respond in kind because the offending operations fall within the

9 Military and Paramilitary Activities in and against Nicaragua (Nicar. v. U.S.), 1986 I.C.J. 14, ¶¶ 176,194 (June 27); Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. 226, ¶ 41 (July 8); Oil Platforms, *supra* note 5, ¶¶43, 76.

10 Tallinn Manual, *supra* note 6.

11 Cyber operations at the use of force level that do not qualify as an armed attack may nevertheless justify countermeasures (see Tallinn Manual).

12 Military and Paramilitary Activities, *supra* note 9, ¶¶ 191 & 210. See also Oil Platforms, *supra* note 5, ¶ 51.

13 Military and Paramilitary Activities, *supra* note 9, ¶ 195.

gap − they are uses of force, but not sufficiently severe to qualify as an armed attack. When this happens the victim-State may resort to either lawful responses, such as diplomatic protests or economic sanctions, or to cyber or kinetic actions short of uses of force that would otherwise be unlawful, but which qualify as lawful "counter measures" in the circumstances.[14] Of course, the victim-State can also refer the matter to the Security Council, which enjoys the authority to act forcefully in the face of any "threat to the peace, breach of the peace, or act of aggression".[15]

Use of the term "armed attack" in lieu of Article 2(4)'s "use of force" verbiage constructs the gap. Note how Article 51 adopts an "act-based" threshold using a specified type of action (armed attack) rather than one based on particular consequences. This approach tracks that taken in Article 2(4), with its prohibition on uses of force. In 1945, an act-based threshold made sense, for the action to which States were most unwilling to completely defer forceful responsive measures to the Charter's new collective security system was an attack by the armed forces of another State. Thus, the term armed attack represented an elegant balancing of the general apprehension about States using force unilaterally, on the one hand, and the fear of States about being defenceless in the face of attacks should the international community fail to act, on the other. This mechanism worked well when the threats that inspired the acceptance of a self-defence exception to the prohibition on the use of force consisted of classic military operations.

The advent of cyber operations challenged this presupposition because dire consequences could now be caused by operations that did not fit neatly into the notion of an attack that was "armed" in the kinetic sense. While the International Court of Justice had opined in its *Nuclear Weapons* advisory opinion that the type of weapon used is immaterial to the application of Articles 2(4) and 51,[16] cyber operations seemed distant from the concept of "armed." Traditional weapons were not employed, they did not require the supporting elements typically associated with military assaults and, most importantly, their direct destructive effect did not result from a release of kinetic force.

The dilemma was that despite these qualitative differences cyber operations could theoretically prove monumentally destructive, in many cases more so than kinetic ones. Accordingly, it was self-evident that some of them were surely encompassed within the ambit of armed attacks. After all, the Charter scheme would make no sense if it prohibited States from responding to devastating attacks merely because such attacks were not in the drafters' contemplation decades before they became technically possible. Such legal formalism would take strict constructionism to absurd ends. Clearly, the advent of cyber operations necessitated a reconceptualization of the notion of "armed attack". To date, the international community has failed to achieve consensus on this critical issue.

The solution to the quandary lies in a realization that the act-based threshold of Article 51 is but cognitive shorthand for a consequence-based legal regime. Reduced to basics, law is about avoiding particular deleterious consequences (or achieving certain positive ones). So the right to resort to force in the face of an armed attack can best be appreciated as a right to do so when States face particular consequences that are severe enough to merit setting aside international

14 On the criteria for, and limitations on, countermeasures, see Articles on State Responsibility, *supra* note 5, ch. 2.

15 U.N. Charter, arts. 39, 42.

16 Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. ¶ 39 (July 8).

law's prohibition on the use of force. By this logic, "armed attack" in the cyber context can be interpreted as encompassing any acts that result in consequences analogous to those caused by the kinetic actions originally envisaged by the term "armed attack."

But what are those consequences? Three points bear on this determinative question. First, as noted, since they are the product of an armed attack, the actions causing them lie above Article 2(4)'s "use of force" threshold. Second, recall the Charter presumption against the use of unilateral force. This too points to a fairly restrictive understanding of armed attack, for it is the point at which States may use force without Security Council authorization. Finally, treaties "shall be interpreted in good faith in accordance with the ordinary meaning to be given to the terms of the treaty in their context and in the light of its object and purpose."[17] The ordinary meanings of the term "armed" are "equipped with or carrying a weapon or weapons," "involving the use of firearms," and "prepared to activate or explode."[18] This suggests that the term implies the sort of consequences that are incident to the use of weapons, an interpretation strengthened by the deliberate omission of the adjective "armed" with respect to "use of force" in Article 2(4). Taken together, a defensible interpretation of the phrase is any action that causes death or injury (including illness and severe suffering) to individuals or damage or destruction of objects.

Some controversy exists over the degree of harm necessary to qualify consequences as an armed attack. The International Court of Justice addressed this matter in the *Nicaragua* case. There it found that an armed attack must have certain "scale and effects," citing the case of a "mere frontier incident" as insufficiently grave.[19] Unfortunately, the Court failed to set forth criteria against which to judge a particular action or incident, an omission for which it has been roundly criticized.[20] In this author's view, it is therefore more useful and appropriate to focus on the qualitative nature of an action's consequences than on any ill-defined quantitative standards; hence the standard proposed.

A recurring question in the cyber context is whether the damage or destruction or manipulation of data that does not generate such consequences is capable of qualifying as an armed attack. Generally it does not, for so qualifying such action would dramatically lower the threshold at which States would enjoy a right to forcefully respond to actions directed at them. This would contravene international law's general presumption against the resort to force in the absence of authorization by the Security Council.

In light of the ever-increasing reliance of society on computers and computer networks, many readers, like the author, will find the "physical consequences" standard too narrow. But it does represent the lex lata, that is, the law that presently exists. For those who share this concern, solace can be found in the fact that international law is not static. As experience with cyber operations grows, the international community may embrace more nuanced understandings of the extant legal standard, or even adopt new legal interpretations thereof. In particular, the law's qualitative focus on the type of harm may yield somewhat to a quantitative analysis such that

17 Vienna Convention on the Law of Treaties art. 31(1), May 23, 1969, 1155 U.N.T.S. 331.

18 The New Oxford American Dictionary, *available at* http://www.oxfordamericandictionary.com/LOGIN?sessionid=35340fb16f7eef9ffa3d1efc76377df8&authstatuscode=400.

19 Military and Paramilitary Activities, *supra* note 9, ¶ 195.

20 And in the later *Platforms* case, it held that the mining of even a single ship could rise to the level of an armed attack. Oil Platforms, *supra* note 5, ¶ 72; see also William H. Taft IV, Self-Defense and the Oil Platforms Decision, 29 Yale Journal of International Law 295, 300 (2004).

a cyber operation causing serious consequences, such as severe economic effects or significant disruption of societal functions, may be characterized as armed attack even if it does not cause death, injury, damage or destruction. Time will tell.

4. CYBER "ATTACKS" UNDER INTERNATIONAL HUMANITARIAN LAW

The notion of armed attacks under the *jus ad bellum* must not be confused with international humanitarian law's usage of the term "attack". In the latter body of law, an "attack" triggers a wide array of legal protections. These prohibitions and restrictions generally derive from the principle of distinction, which requires the parties to a conflict to "at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly direct their operations only against military objectives."[21]

Although the principle of distinction is framed in terms of "military operations," it is clear that not all military operations are contemplated by the norm. For instance, longstanding State practice demonstrates that non-destructive psychological operations directed at the civilian population, such as dropping leaflets, broadcasting to the enemy population, or even jamming enemy public broadcasts, are lawful as long as no physical consequences attend them. Rather, the principle is primarily meant to address "attacks", as that term is understood in the law.

Various facts support this contention. Note how the principle of distinction is set forth in Article 48 of Additional Protocol I. That article appears in the Chapter on "Basic Rule and Field of Application" of the treaty's conduct of hostilities section. Since the only other article in the Chapter is Article 49, which defines attacks, this placement implies that the military operations referred to in Article 48 are primarily attacks.

Further review of the section reveals a constant and pervasive emphasis on "attacks". Article 51 is illustrative. It begins by noting that the "civilian population and individual civilians shall enjoy general protection against dangers arising from military *operations*," but operationalizes the provision by noting that "to give effect to this protection" it is prohibited to attack individual civilians or the civilian population, conduct an attack that is not directed at a military objective, engage in reprisal attacks against civilians, launch attacks in which the expected collateral damage is excessive relative to anticipated military advantage, treat multiple military objectives during an attack as a single one when they are clearly separated and distinct in a concentration of civilians, and use a method or means of warfare during an attack that is either incapable of distinguishing lawful from unlawful targets or has effects that cannot be controlled.[22] Subsequent articles are likewise framed in terms of prohibitions and restrictions on attacks. The most important of these prohibit attacks on civilian objects and mandate various precautions that must be taken during an attack to avoid harming the civilian population and civilians. Simply put, the prohibition on directing military operations against civilians, civilian objects

21 AP I, *supra* note 7, art. 48. The provision is generally deemed reflective of customary international law and the International Court of Justice has cited it as one of international humanitarian law's "cardinal" principles. Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. 226, ¶ 78 (July 8).

22 AP I, *supra* note 7, art. 51(4). The emphasis in this and all other treaty extracts is the author's and does not appear in the original.

and other protected persons and objects must be understood as essentially a prohibition on *attacking* them. Conducting military operations that do not qualify as attacks against them is, in a general sense, lawful (absent a specific prohibition to the contrary[23]).

This conclusion raises the question of which acts qualify as an attack. The reference to acts of violence against the adversary, whether in offence or defence, in Article 49 is the key to the answer.[24] It should be cautioned that mention of the "adversary" does not imply that only violent operations against enemy forces qualify. On the contrary, the prohibition on attacking civilians irrefutably confirms that the *sine qua non* criterion is violence, not the individual or entity that is the object of an attack.

The definitional centrality of violence is well supported. For example, the Bothe, Partsch and Solf commentary on Additional Protocol I explains that "[t]he term 'acts of violence' denotes physical force. Thus, the concept of 'attacks' does not include dissemination of propaganda, embargoes, or other non-physical means of psychological or economic warfare."[25] Their commentary is particularly authoritative given that all three were active participants at the Diplomatic Conference that negotiated the treaty. The official International Committee of the Red Cross (ICRC) Commentary similarly explains that "the term 'attack' means 'combat action.'"[26]

The cognitive dilemma is that cyber operations do not directly involve the release of violent forces. This begs the questions of whether and when cyber operations qualify as attacks under international humanitarian law such that its prohibitions and restrictions thereon apply.

As with the UN Charter, actions that can cause harm without the immediate release of violent kinetic forces were beyond the contemplation of the drafters of Additional Protocol in 1977. Yet, by then, an implicit recognition existed that the violence of an act itself was not the crux of the norms in question. Over a half-century earlier, employment of chemical and biological weapons was already considered an attack, as evidenced, *inter alia*, by the outlawing of their use for Parties to the 1925 Gas Protocol.[27] They were outlawed because they were instrumentalities that caused particular harmful consequences that international humanitarian law sought to avoid. By the same logic, "acts of violence" are merely instrumentalities that cause consequences with which the law concerns itself.

Moreover, as noted, treaties must be interpreted in "context and in light of object and purpose." A careful reading of Additional Protocol I's prohibitions and restrictions on attacks discloses that the concern was not so much with acts which were violent, but rather with those that have harmful consequences (or risk them), in other words, violent consequences. In great part, the treaty's object and purpose is to avoid, to the extent possible in light of military necessity, those very consequences. For instance, civilians "enjoy general protection against *dangers* arising

23 As with the requirement to "respect and protect" medical units in addition to the prohibition on attacking them. AP I, *supra* note 7, art. 12.

24 See text accompanying note 7.

25 Michael Bothe et al., New Rules for Victims of Armed Conflicts 289 (1982).

26 Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949, ¶ 1880 (Yves Sandoz, Christophe Swinarski & Bruno Zimmermann, eds., 1987)

27 1925 Geneva Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare, June 17, 1925, 26 U.S.T. 571, T.I.A.S. No. 8061.

from military operations."[28] Acts intended to *terrorize* the civilian population are prohibited.[29] The rule of proportionality assesses an act in light of the "incidental *loss* of civilian life, *injury* to civilians, *damage* to civilian objects, or a combination thereof" expected to be caused by an attack.[30] Precautions that are required to be taken when conducting an attack are meant to "*spare*" the civilian population.[31] They include selecting weapons and tactics "with a view to avoiding, and in any event to minimizing, incidental *loss* of civilian life, *injury* to civilians and *damage* to civilian objects"; refraining from launching, suspending, and cancelling attacks that would likely cause excessive "incidental *loss* of civilian life, *injury* to civilians [or] *damage*"; issuing warnings when feasible if an attack will "*affect* the civilian population"; choosing among comparable targets "which may be expected to cause the least *danger* to civilian lives and to civilian objects"; and, in air and sea operations, taking precautions "to avoid *losses* of civilian lives and *damage* to civilian objects".[32] Defenders must similarly take measures to protect civilians and civilian objects from "*danger*".[33] The same consequence-based approach applies to specially protected objects, as in the restrictions on conducting attacks against dams, dykes and nuclear generating stations when "*severe losses*" among the civilian population might result[34] and the prohibition on using methods or means of warfare likely to cause "*widespread, long-term and severe damage*" to the natural environment and thereby "prejudice the *health or survival* of the population."[35]

It is apparent that international humanitarian law, despite adopting an instrumentality-based definition of attack, takes a consequence-based approach to it normative prescriptions when operationalizing that term. The Bothe, Partsch and Solf commentary to Article 49 supports this conclusion by noting that attack refers to "those aspects of military operations that *most directly affect* the safety of the civilian population and the integrity of civilian objects."[36]

Through the process of induction, it is possible to derive a general principle regarding the notion of attack that has meaning within the cyber context. Attacks can be redefined as operations that result in, or if unsuccessful were originally expected to result in, death or injury of individuals or destruction or damage of objects. The notion of injury includes illness that might result from a cyber operation, as in the case of attacking a water treatment plant in order to contaminate drinking water. It is also sensible, based for example on the prohibition of terror attacks and starvation[37], to extend the concept to acts producing serious suffering not otherwise justified by the notion of military necessity. Destruction includes operations that, while not causing physical damage, nevertheless "break" an object, rendering it inoperable, as in the case of a cyber operation that causes a computer reliant system to no longer function unless repaired. Thus, the legal analysis of attack in the international humanitarian law context leads to roughly the same conclusion as arrived at with respect to the *jus ad bellum*. However, the reader must understand that since they derive from different bodies of law, their precise parameters are nuanced in ways beyond the capability of this article to address.[38]

28 AP I, *supra* note 7, arts. 51(2).
29 *Id*., art. 51(3).
30 *Id*., arts. 51(5)(b) & 57(2)(a)(iii).
31 *Id*., art. 57(1).
32 *Id*., art. 57.
33 *Id*., art. 58.
34 *Id*., art. 56(1).
35 *Id*., art. 55(1).
36 Bothe, *supra* note 26, at 325.
37 AP I, *supra* note 7, arts. 51(2) & 54.
38 These nuanced are explored in the forthcoming Tallinn Manual, *supra* note 6.

The consequence of this conclusion for cyber operations is significant. It means that cyber operations can be directed at civilian systems so long as the requisite type of harm is not triggered and no other specific international humanitarian law prohibition (such as those attending medical operations) applies.

At the 37th International Conference of the Red Cross and Red Crescent Society in 2011, the ICRC circulated a background paper articulating a different approach.[39] It began by noting that Article 49's reference to "acts of violence […] denotes physical force." Accordingly, "cyber operations by means of viruses, worms, etc., that result in physical damage to persons, or damage to objects that goes beyond the computer program or data attacked could be qualified as 'acts of violence', i.e. as an attack in the sense of IHL." There is universal agreement on this point.

However, the document then took issue with the general approach set forth (except for reversibility) in this article, that is, that "cyber operations do not fall within the definition of 'attack' as long as they do not result in physical destruction or when its effects are reversible." According to the ICRC paper,

> "[i]f this claim implies that an attack against a civilian object may be considered lawful in such cases, it is unfounded under existing law in the view of the ICRC. Under IHL, attacks may only be directed at military objectives, while objects not falling within that definition are civilian and may not be attacked. The definition of military objectives is not dependent on the method of warfare used and must be applied to both kinetic and non-kinetic means; the fact that a cyber operation does not lead to the destruction of an attacked object is also irrelevant. Pursuant to article 52 (2) of Additional Protocol I, only objects that make an effective contribution to military action and whose total or partial destruction, capture or neutralization offers a definite military advantage, may be attacked. By referring not only to destruction or capture of the object but also to its neutralization the definition implies that it is immaterial whether an object is disabled through destruction or in any other way."[40]

The ICRC's references to international humanitarian law comments reflect the state of the law. There is no doubt that an attack against a civilian object is unlawful. Nor is there any doubt that the methods or means of attack have no bearing whatsoever on the legal character of a targeted object as either a civilian object or a military objective. And the reference to "neutralization" properly confirms that the military advantage required for qualification as a military objective need not stem from physical damage to the target. These are binding norms not only for Parties to Additional Protocol I, but for also for other Stats since they reflect customary international law.[41]

But the organization's conclusion misses the mark. The question at hand is whether a cyber operation qualifies as an attack in the first place. Only when it does is the issue of the target's

39 International Committee of the Red Cross, 31st International Conference of the Red Cross and Red Crescent, International Humanitarian Law and the Challenges of Contemporary Armed Conflicts, Report 31IC/11/5.1.2, Oct. 2011.

40 *Id*. at 37.

41 See, e.g., Department of the Navy et al., The Commander's Handbook on the Law of Naval Operations (NWP 1-14M/MCWP 5-12/COMDTPUB P5800.7A), chapter 8 (2007).

status raised, for only then do international humanitarian law prohibitions and restrictions as to attacks come into play. Consequently, once a cyber operation qualifies as an attack, Article 52(2)'s criteria for qualification as a military objective apply…and not before that determination is made. Should an object not constitute a military objective, a prospective attack thereon is prohibited. If it does, the object may, as a military objective, be attacked by any method or means of warfare that otherwise complies with the rule of proportionality, the requirement to take precautions in attack and other applicable standards. For instance, even when cyber operation can be employed to neutralize a military objective, an attacker may elect to bomb it doing so is not expected to exacerbate incidental harm to civilians, civilian objects and other protected persons and places.

Admittedly, the conclusions reached in this article regarding the meaning of "attack" in international humanitarian law may seem unsatisfactory. Non-destructive attacks and those that do not place individuals or objects at physical risk can have severe consequences. Yet, the interpretation advanced in this article represents the extant law, that is, the *lex lata*. Assertions to the contrary are, in the author's estimation, merely *lex ferenda*. Of course, as with the term "armed attack" in the *jus ad bellum* context, the meaning of a legal term may shift over time through adoption of new treaty law, creation of new customary norms through State practice, or the emergence of new understandings in the face of the changing context of conflict to which it applies.

5. CONCLUDING THOUGHT

This article has attempted to clear some of the terminological dissonance that exists between the policy/technical/operational and legal communities regarding the term "attack." The former must be sensitive to the fact that legal meaning also attaches when the term is used in its colloquial sense. Complicating matters is the fact that the term inhabits two separate and distinct areas of the law. The risk of creating confusion as to precise policy parameters is accordingly high when using the term without care. For its part, the legal community must be alert to the possibility that its legal advice may not be fully grasped by their clients when the term attack is used *stricto sensu*. Unfortunately, the dearth of systematic interaction between the respective cyber communities has resulted in the emergence of two patois that are sometimes unintelligible to each other. It is hoped that this book, and the conference upon which it is based, will serve to narrow the gap between them.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Ius ad bellum in Cyberspace – Some Thoughts on the "Schmitt-Criteria" for Use of Force

Katharina Ziolkowski
Legal & Policy Branch
NATO CCD COE
Tallinn, Estonia
katharina.ziolkowski@ccdcoe.org

Abstract: The term "cyber-attack" has become a synonym for any malicious cyber-activity. Given the martial semantics and the hype of "cyberwar" in the media and non-legal disciplines, as well as the political sabre-rattling partly perceivable in international relations, the present article endeavours to augment the academic discussion in regard to the criteria used for the legal assessment of malicious cyber-activities as "use of force" pursuant to Article 2(4) of the UN Charter and, at the same time, in regard to the closely related term of "armed attack" in the meaning of Article 51 of the UN-Charter and Article 5 of the North Atlantic Treaty. The importance of the discussion of such criteria lies in the fact that "use of force" in international relations enables the victim State to undertake a range of unfriendly (retorsions) and otherwise illegal actions (counter-measures), and that an "armed attack" triggers the right to self-defence of the victim State and justifies its resort to forceful self-defence measures – all situations with potentially severe consequences for international peace and security. First, the traditional meaning of the terms "use of force" and "armed attack" will be presented. Without replicating the relevant scholarly writings, it will be shown which categories of malicious cyber-activities can be considered "use of [armed] force" and – given a certain threshold of severity in scale and effects – as an "armed attack". In this context, the so-called "Schmitt-Criteria" for the classification of malicious cyber-activities as "use of force", established by Professor Michael N. Schmitt over a decade ago and hitherto not analysed in depth within scholarly writings, will be elaborated upon. These criteria contain a range of significant aspects and refer to complex matters; therefore, they deserve a substantial discussion. Due to the focus and the limited scope of the present paper, the discussion of the *ius ad bellum* aspects related to Chapter VI and VII of the UN Charter will be deliberately omitted.

Keywords: *cyberspace, use of force, armed attack, Art. 2(4) UN Charter, Art. 51 UN Charter, Art. 5 North Atlantic Treaty, Schmitt-Criteria*

1. INTRODUCTION

Since the term "cyber-attack" has become a synonym for any malevolent activity conducted by the means of the Internet or other information and communication technologies (in the following referred to as "malicious cyber-activity"), a martial connotation can be perceived in the respective semantics describing cyber-threats and malicious cyber-activities. Especially media and non-legal disciplines use the term "attack" without the necessary sensitivity, which would be desirable, given the cognitive association of the term in the context of international peace and security. Additionally, the different meanings of the legal term "attack" being a term of art for both, the *ius ad bellum* and in the *ius in bello*, are not always clearly distinguished[1].

Given the confusion in terminology, and bearing in mind the aforementioned martial semantics, the media-hype of "cyberwar" and the political sabre-rattling partly perceivable in international relations[2] – clearly to be seen in the context of deterrence policy efforts –, the present article endeavours to augment the academic discussion in regard to the criteria used for the legal assessment of malicious cyber-activities as "use of force" pursuant to Article 2(4) of the UN Charter, enabling States to undertake a range of unfriendly (retorsions) and otherwise illegal actions (counter-measures), and in regard to the closely related term of an "armed attack", justifying a State's resort to self-defence measures in the meaning of Article 51 of the UN Charter and Article 5 of the North Atlantic Treaty. In particular, the so-called "Schmitt-Criteria" for the classification of malicious cyber-activities as "use of force", established by Professor Michael N. Schmitt over a decade ago and – pursuant to the knowledge of the author – hitherto not analysed in depth within scholarly writings, will be elaborated upon. The criteria contain a range of significant aspects and refer to complex matters; therefore, they deserve a substantial discussion. The assessment will, *inter alia*, show differences in the approach of the common law system and the civil law system in regard to lines of legal argumentation.

However, it shall be emphasised that the decision about undertaking retorsions or counter-measures, as well as about the existence of a self-defence situation and the resort to use of force in international relations will always be a political one, which will be taken at the highest levels of a State's governmental structure and which will always depend on the overall political context of the particular political crisis. The legal discipline can only support governmental decision-makers by providing in advance abstract criteria and – in the case of governmental legal advisors – concrete *ad hoc* legal counsel affecting the overall assessment and judgment.

It shall be only mentioned that, due to the focus and the limited scope of the present survey, the discussion of the *ius ad bellum* aspects related to Chapter VI and VII of the UN Charter is deliberately omitted.

1 *See* M. N. Schmitt, "'Attack' as a Term of Art in International Law: The Cyber Operations Context" in this volume.

2 *See e.g.* S. Gorman & J. E. Barnes, "Cyber Combat: Act of War Pentagon Sets Stage for U.S. to Respond to Computer Sabotage With Military Force", in *The Wall Street Journal* online of 31 May 2011, *available at* http://online.wsj.com/article/SB10001424052702304563104576355623135782718.html (last visited 16 April 2012).

2. "USE OF FORCE" AND "ARMED ATTACK" IN PUBLIC INTERNATIONAL LAW

Currently, neither a legal definition nor a universally accepted definition of the terms "use of force" (Article 2(4) of the UN Charter) and "armed attack" (Article 51 of the UN Charter, Article 5 of the North Atlantis Treaty) exist. However, the meaning of the terms can be clarified to a certain degree by substantial interpretive work, an endeavour challenged by the fact that the core meanings of the treaty norms are recognised to constitute norms of international customary law at the same time.

As indicated by Articles 31-33 of the Vienna Convention on the Law of Treaties and the corresponding[3] international customary law, and by Article 38(1) of the Statute of the International Court of Justice (ICJ), the interpretation of a term should include, *inter alia*, the preparatory work of the treaty and the ordinary meaning of the term in its context of the treaty and in the light of its object and purpose. These aspects reflect the canon of legal interpretation, stated by the German lawyer Friedrich Carl von Savigny[4] in the early 19th century and still forming an elementary component of legal teaching in continental-Europe: the historic, the textual, the systematic and the teleological interpretation. Corresponding to the nature of public international law, the aforementioned norms designate further aspects to be taken into account when interpreting international norms. Those are, among others, subsequent State practice or international custom, judicial decisions and, according to Article 38(1)(d) of the ICJ-Statute, "the teachings of the most highly qualified publicists of the various nations". It shall be mentioned that in regard to *ius ad bellum* as applicable to cyberspace, it is the work of academia which currently importantly influences the development of a common understanding within the international community. Potential State practice is not perceivable in the public, declarations of *opinio iuris* by States are rare and general[5] in nature, and respective national or international jurisdiction does not yet exist on the matter.

In the following, first, the traditional meaning of the terms will be presented, before its application to acts conducted by means of the Internet or other information and communication technologies will be elaborated upon.

Although disputed in detail, it can be stated that – generally speaking – an "armed attack" is given in most severe cases of "use of force" in international relations (in the meaning of Article 2(4) of the UN-Charter) of significant scale and effects. This finding is supported by the

3 Despite being highly political documents, the UN Charter and the North Atlantic Treaty are subject to the rules of interpretation of international treaties. Although, according to its Article 4, the Convention of 1969 does not apply retroactively (to the UN Charter of 1945 and to the North Atlantic Treaty of 1949), the provisions on interpretation of treaties are a valuable reference as they reflect international customary law. *See*: G. Ress, "The Interpretation of the Charter", in B. Simma (ed.), *The Charter of the United Nations. A Commentary* (Oxford / New York, Oxford University Press, 2002, 2nd ed.), at para. 2 et seq.; ICJ, *Oil Platforms (Islamic Republic of Iran v. United States of America)*, Merits, ICJ Rep. 1996, at p. 823 para. 41.

4 For more information on von Savigny see "Friedrich Karl von Savigny", in *Encyclopedia Britannica Online, available at* http://www.britannica.com/EBchecked/topic/525746/Friedrich-Karl-von-Savigny (last visited 16 April 2012).

5 *See* Gorman & Barnes, *supra* note 2.

jurisdiction of the ICJ[6] as well as by a vast amount of scholarly writings[7], of which the mere repetition will be abstained from in this survey.

Thus, the two terms "use of force" and "armed attack" are closely related. In order to identify which situations would comprise an "armed attack" and trigger the right of the victim State to undertake self-defence measures it must first be established in which situations "force" in the meaning of Article 2(4) of the UN Charter is used in international relations. Thus, the term "use of force" can be deemed as the nucleus of all *ius ad bellum* deliberations.

Illustrating the different lines of arguments concerning a further specification of the term "force" within international jurisdiction and scholarly writings would certainly exceed the scope of this paper. In addition, there is no benefit in their mere replication. Therefore, without further explanation, in the following it will be assumed that "force" in the meaning of Article 2(4) of the UN Charter is to be understood as "armed force".[8] Hereby, two aspects are of importance for further deliberations: On the one hand, pursuant to the historical, systematic and teleological interpretation of the norm, "use of [armed] force" does not include measures of mere coercion, be it political or economic in nature.[9] On the other hand, however, the term "use of [armed] force" is not limited to the employment of military weaponry: The ICJ stated over 25 years ago the possibility of an "indirect" use of armed force[10] (e.g. by arming and training insurgents) and scholarly writings[11] describe e.g. spreading fire over the border or flooding another State's territory as violating the prohibition of "use of [armed] force".

In order to specify the meaning of "use of [armed] force" conducted by the means of the Internet or other information and communication technologies, an effects-based approach inherent to public international law is surely to be considered appropriate (ruling out other possible approaches, e.g. focusing the target of the malicious activities, the intent of the malevolent actor, or the designation of the means used). Hereby, the comparison of the effects indirectly caused or intended by malicious cyber-activities with the effects usually caused or intended by conventional, biological or chemical weapons (BC-weapons) plays a paramount role[12].

Again, in order not to replicate the legal arguments presented in diverse scholarly writings, it

6 The ICJ held in the *Nicaragua Case* that only "the most grave forms" of use of force "[…] of significant scale […]", which "[…] because of its scale and effects, would have been classified as an armed attack rather than a mere frontier incident […]" could trigger the right to self-defence; see ICJ, *Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v. United States of America)*, Merits, ICJ Rep. 1986, pp. 14–150, at pp. 101 and 103 paras. 191 and 195; *see* also ICJ, *Oil Platforms (Islamic Republic of Iran v. United States of America)*, ICJ Rep. 2003, at p. 161 para. 51.

7 *See* A. Randelzhofer, "Article 51", in B. Simma (ed.), supra note 3, at paras. 4 and 20; M. Bothe, "Völkerrechtliche Verhinderung von Gewalt *(ius contra bellum)*", in W. Graf Vitzthum, *Völkerrecht* (Berlin, De Gruyter, 2001), Section 8, at para. 10; R. Higgins, *Problems and Process: International Law and How We Use It* (Oxford, Oxford University Press, 1994), at p. 250.

8 A good overview on the discussion is given by M. Roscini, "Word Wide Warfare – Jus ad bellum and the Use of Cyber Force", Vol. 14 *Max Planck Yearbook of United Nations Law* 2010, pp. 85-130, at pp. 104-106; *see* also A. Randelzhofer, "Article 2(4)", in B. Simma (ed.), *supra* note 3; Th. Bruha, "Use of Force, Prohibition of", in R. Wolfrum & Ch. Philipp (eds.), *United Nations: Law, Policies and Practice*, (Vol. II., München, Springer, 1995), at pp. 1387 *et seq*.

9 Randelzhofer, supra note 8, at para. 21.

10 ICJ, *Case Concerning Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America)*, Merits, ICJ Rep. 1986, pp. 14–150, at p. 118 *et seq*. para. 228.

11 *See* e.g. Randelzhofer, *supra* note 8.

12 For detailed discussion see K. Ziolkowski, "Computer Network Operations and the Law of Armed Conflict", Vol. 49 *Military Law and the Law of War Review* 2010, pp. 47-94, at pp. 69-75.

can be assumed that malicious cyber activities can be considered "use of [armed] force" in the meaning of Article 2(4) of the UN Charter if they – indirectly – result in[13]:

- Deaths or physical injuries of living beings and/or the destruction of property.[14]
- Massive, medium to long-term disruption of critical infrastructure systems of a State (if in its effects equal to the physical destruction of the respective systems).[15]

Neither the destruction of data (even of substantial importance, e.g. classified data, or of significant economical value, e.g. symbolising assets)[16] nor the "theft"[17] (rather: illegal copying) of data (being nothing more than modern espionage[18] not generally forbidden under public international law) can be considered "use of [armed] force".[19] Such effects cannot be equated to the effects usually caused or intended by conventional or BC-weapons, especially not to the physical destruction of objects.[20] Furthermore, it is agreed by the vast majority of scholars, that malicious cyber-activities targeted at critical infrastructure systems of a State, which do not exceed the threshold of merely minimally affecting the population's quality of life or going beyond a mere inconvenience, are not showing effects of disruption of the public life

13 *Ibid*.

14 Y. Dinstein, "Computer Network Attack and Self-Defense", in M.N. Schmitt & B.T. O'Donnell (eds.), *Computer Network Attack and International Law* (Newport / Rhode Island, US Naval War College, 2002), pp. 59–71, at p. 103; D.B. Silver, "Computer Network Attack as a Use of Force under Article 2(4) of the United Nations Charter", in Schmitt & O'Donnell, at p. 85; J. Barkham, "Information Warfare and International Law on the Use of Force", Vol. 34 *New York University Journal of International Law & Politics* 2001, at p. 80; T. Morth, "Considering Our Position. Viewing Information Warfare as Use of Force Prohibited by Article 2(4) of the U.N. Charter", Vol. 30 *Case Western Reserve Journal of International Law* 1998, at p. 591; T. Stein & T. Marauhn, "Völkerrechtliche Aspekte von Informationsoperationen", Vol. 60 *Zeitschrift für ausländisches öffentliches Recht und Völkerrecht* 2000, pp. 1-60, at p. 7; C.C. Joyner & C. Lotrionte, "Information Warfare as International Coercion: Elements of a Legal Framework", Vol. 12 *European Journal of International Law* 2001, at pp. 846 and 850; M.N. Schmitt, "Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework", Vol. 37 No. 3 *Columbia Journal of Transnational Law* 1999, at pp. 914 *et seq*.; W.G. Sharp, *Cyberspace and the Use of Force* (Falls Church, Aegis Research Cooperation, 1999), at p. 102; and L.T. Greenberg, S.E. Goodman & K.J. Soo Hoo, *Information Warfare and International Law* (Washington, National Defence University, 1998), at pp. 19 and 32.

15 Ziolkowski, *supra* note 12, at pp. 69-75; J.P. Terry, "Responding to Attacks on Critical Computer Infrastructure. What Targets? What Rules of Engagement?", in Schmitt & O'Donnell (eds.), *supra* note 14, at pp. 428 *et seq*.; Morth, *supra* note 14, at p. 599; Sharp, *supra* note 14, at pp. 129 *et seq*. *Contra:* Dinstein, *supra* note 14, at p. 105; and Stein & Marauhn, *supra* note 14, at p. 8, who demand the occurrence of physical damage outside the targeted computer networks in order to qualify CNO as use of force.

16 *See* Barkham, *supra* note 14, at p. 88.; M.N. Schmitt, D.H.A. Harrison & Th.C. Wingfield, *Computers and War: The Legal Battlespace* (International Humanitarian Law Research Institute, Background Paper, 2004), at pp. 5 *et seq*.

17 Joyner & Lotrionte, *supra* note 14, at pp. 846, 855 *et seq*.; *contra:* Stein & Marauhn, *supra* note 14, at p. 10.

18 A. D'Amato, "International Law, Cybernetics, and Cyberspace", in M.N. Schmitt & B.T. O'Donnell (eds.), *supra* note 14, pp. 59–71, at p. 67; and Stein & Marauhn, *supra* note 14, at p. 32 with further references. In regard to cyber-activities as a modern form of espionage see W.H. von Heinegg, "Informationskrieg und Völkerrecht. Angriffe auf Computernetzwerke in der Grauzone zwischen nachweisbarem Recht und rechtspolitischer Forderung", in V. Epping , H. Fischer & W.H. von Heinegg (Hrsg.), *Brücken bauen und begehen. Festschrift für Knut Ipsen zum 65. Geburtstag* (München, C.H. Beck, 2000), at p. 134. Apart from the penalisation of espionage resulting from respective national law systems, spying is restrained by certain provisions of public international law, e.g. the taboos stated by the diplomatic and consular law protecting diplomatic and consular archives and correspondence, i.e. respective electronic databases and communication via the Internet.

19 Ziolkowski, *supra* note 12, at pp. 69-75.

20 For detailed discussion *see ibid*.

and *ordre public* similar to physical destruction by e.g. a bombardment and, therefore, do not amount to "use of [armed] force".

3. THE "SCHMITT-CRITERIA"

"Use of [armed] force" in the meaning of Article 2(4) of the UN Charter is to be distinguished especially from measures of mere (economic or political) coercion[21] in international relations, a task that can pose considerable challenges upon decision-makers in practice. For facilitating such a distinction, in 1999[22] Professor Schmitt developed and recently reinforced[23] a set of criteria for the determination of "use of [armed] force" (amending their descriptions over time). The factors shall serve as indicators which States are likely to take into consideration when assessing whether specific malicious cyber-activities qualify as "use of [armed] force".[24]

These criteria are:[25]

> "1) *Severity:* Consequences involving physical harm to individuals or property will alone amount to a use of force. Those generating only minor inconvenience or irritation will never do so. Between the extremes, the more consequences impinge on critical national interests, the more they will contribute to the depiction of a cyber operation as a use of force. In this regard, the scale, scope, and duration of the consequences will have great bearing on the appraisal of their severity. Severity is self-evidently the most significant factor in the analysis.
>
> 2) *Immediacy:* The sooner consequences manifest, the less opportunity states have to seek peaceful accommodation of a dispute or to otherwise forestall their harmful effects. Therefore, states harbor a greater concern about immediate consequences than those that are delayed or build slowly over time.
>
> 3) *Directness:* The greater the attenuation between the initial act and the resulting consequences, the less likely states will be to deem the actor responsible for violating the prohibition on the use of force. Whereas the immediacy factor focused on the temporal aspect of the consequences in question, directness examines the chain of causation. For instance, the eventual consequences of economic coercion (economic downturn) are determined by market forces, access to markets, and so forth. The causal connection between the initial acts and their effects tends to be indirect. In armed actions, by contrast, cause and effect are closely related—an explosion, for example, directly harms people or objects.
>
> 4) *Invasiveness:* The more secure a targeted system, the greater the concern as to its penetration. By way of illustration, economic coercion may involve no intrusion at all

21 See representatively: Randelzhofer, *supra* note 8, at para. 21.
22 Schmitt, *supra* note 14, at pp. 913 *et seq*.
23 M.N. Schmitt, "Cyber Operations and the *Jus Ad Bellum* Revised", Vol. 56 *Villanova Law Review* 2011, at pp. 576 *et seq*. The criterion of "responsibility" was mentioned already in the 1999 publication, although only in a footnote, see Schmitt, *supra* note 14, at p. 915, footnote 81.
24 *Id*., at p. 605.
25 *Id*., at pp. 576 *et seq*.

(trade with the target state is simply cut off), whereas in combat the forces of one state cross into another in violation of its sovereignty. The former is undeniably not a use of force, whereas the latter always qualifies as such (absent legal justification, such as evacuation of nationals abroad during times of unrest). In the cyber context, this factor must be cautiously applied. In particular, cyber exploitation is a pervasive tool of modern espionage. Although highly invasive, espionage does not constitute a use of force (or armed attack) under international law absent a nonconsensual physical penetration of the target state's territory, as in the case of a warship or military aircraft which collects intelligence from within its territorial sea or airspace. Thus, actions such as disabling cyber security mechanisms to monitor keystrokes would, despite their invasiveness, be unlikely to be seen as a use of force.

5) *Measurability:* The more quantifiable and identifiable a set of consequences, the more a state's interest will be deemed to have been affected. On the one hand, international law does not view economic coercion as a use of force even though it may cause significant suffering. On the other, a military attack that causes only a limited degree of destruction clearly qualifies. It is difficult to identify or quantify the harm caused by the former (e.g., economic opportunity costs), while doing so is straightforward in the latter (X deaths, Y buildings destroyed, etc).

6) *Presumptive legitimacy:* At the risk of oversimplification, international law is generally prohibitory in nature. In other words, acts which are not forbidden are permitted; absent an express prohibition, an act is presumptively legitimate.[...] For instance, it is well accepted that the international law governing the use of force does not prohibit propaganda, psychological warfare, or espionage. To the extent such activities are conducted through cyber operations, they are presumptively legitimate.

7) *Responsibility:* The law of state responsibility [...] governs when a state will be responsible for cyber operations. But it must be understood that responsibility lies along a continuum from operations conducted by a state itself to those in which it is merely involved in some fashion. The closer the nexus between a state and the operations, the more likely other states will be inclined to characterize them as uses of force, for the greater the risk posed to international stability."

4. SOME THOUGHTS ON THE "SCHMITT-CRITERIA"

The criteria, which – pursuant to the knowledge of the author – hitherto have not been analysed in depth within academic writings, contain a range of significant aspects and refer to complex matters; therefore, they deserve a substantial discussion. The following considerations aim to initiate such a debate.

Severity

As Professor Schmitt states, the criterion of "severity" is the most significant in the analysis of malicious cyber-activities. Insofar as the criterion refers to "physical harm to individuals or property", it is congruent with the above presented view that malicious cyber activities indirectly

resulting in "deaths or physical injuries of living beings and/or the destruction of property" can be considered "use of [armed] force" in the meaning of Article 2(4) of the UN Charter. The author of the present survey would argue that the "massive, medium to long-term disruption of critical infrastructure systems of a State (if in its effects equal to the physical destruction of the respective systems)" would also be covered by the "Schmitt-Criterion" of "severity". Disabling critical infrastructure systems, massive in scope and duration, can be equated to "physical harm to property" in the sense of eliminating the functionality of the targeted systems. In either the case of kinetic destruction of the components of a critical infrastructure system or in the case of total disabling of the system, the system in question cannot serve its purpose and must be – in whatever way – repaired in order to function.

The author of the present survey subscribes to the criterion of "severity" and its importance, except for the aspect of the relevance of "critical national interests" ("Between the extremes, the more consequences impinge on critical national interests, the more they will contribute to the depiction of a cyber operation as a use of force"). The prohibition of the "use of [armed] force" in international relations in the meaning of Article 2(4) of the UN Charter (and the right to self-defence, given in most severe cases of "use of [armed] force") does not protect the national interests – which can be manifold, including e.g. economic interests – but rather the (physical) security of a State and its population. This threshold is high and, for the sake of international peace and security, should not be diluted.

As a final thought on the criterion of "severity" of the effects of malicious cyber-activities, the author of the present survey argues that in the future a debate on the so-called "accumulation of events" or "Nadelstichtaktik" doctrine will present a necessary part of the discussion in regard to cyberspace. The abovementioned approaches were elaborated in the legal literature in order to categorise the "hit and run" or guerrilla tactics within the *ius ad bellum* and where used in the political practice of the USA and Israel in the course of justifications of forceful measures conducted against "terrorists" in the past (partly condemned by the UN Security Council as "retaliation").[26] This thought is based on a certain tendency visible in cyberspace. Malevolent data-streams, accumulating to a malicious code at its destination, are being deliberately sent in an extremely slow manner and in small pieces in order to be classified by the security-sensors of the targeted computer systems as "background noise" and not as a danger. It is conceivable that in the future such a segmented course of action could also be conducted in regard to the (physical) effects caused by malicious cyber-activities. For example, the malfunctioning of a few critical infrastructure systems of a State could be caused by and by, each of which would rather be classified as a mere nuisance than "use of [armed] force" – a finding which could turn out differently if the malfunctioning of the different systems at different times were judged in terms of their "accumulation".

Immediacy

The explanatory text to the criterion suggests that "immediacy" of consequences of malicious cyber-activities is an aspect but not a requirement for their classification as "use of [armed] force".

26 *See* examples of State practice, UN Security Council resolutions and a discussion in K. Ziolkowski, *Gerechtigkeitspostulate als Rechtfertigung von Kriegen* (Baden-Baden, Nomos, 2008), at pp. 229-231.

This is of importance because it is very likely that the consequences of malicious cyber-activities – even if immediately given – will mostly not be recognisable or not recognised as such for a certain period of time. This is based on the complexity of modern computer systems and the large number of possible errors, which can lead to the malfunctioning of the systems. In the case of malfunctioning of computer systems it will always be investigated first whether the problem is caused by a programming error of the software-producer, by a malfunction of outsourced computer services providers, by mal-configuration of the systems by the own system administrators, or by errors of the users of the system. Additionally, it is conceivable that in cases of malicious cyber-activities against critical infrastructure systems of a State, a vast majority of which is owned and operated by private industry, both intrusions into the computer systems and their perceptible effects would be covered in order to not lose confidence in the security of the respective services and to preserve the own reputation and the customers' trust. A long period of time can pass by before malicious data-streams will be discovered and analysed and finally brought into context with the negative effects on governmental levels dealing with questions of national security and foreign policy. For example, the worm Stuxnet was discovered in July 2010 in the computer systems of Iranian nuclear power installations, "but is confirmed to have existed at least one year prior and likely even before"[27]. By February 2010 the IT-security company *Symantec* – monitoring the command and control traffic of the worm – had gathered 3,280 unique samples representing three different variants of *Stuxnet*.[28] Media reports of the replacement[29] of a remarkable number of centrifuges in the nuclear enrichment facility at *Natanz* could – although hitherto not confirmed[30] by Iranian officials – indicate that the effects of the malicious codes were conceivable in the past but not brought into context with a possible computer system problem. However, as e.g. border intrusions by military forces of a neighbouring State in a (geographically) remote area of a victim State's territory would constitute a "use of [armed] force", although not recognisable to the victim State immediately, malicious cyber-activities, although their perceptible (physical) effects are not recognisable yet as such, can also theoretically be classified as "use of [armed] force".

Thus, given the complexity of cyberspace and the large number of possible reasons for malfunctioning of computer systems, the recognition of the connection between malicious cyber-activities and their perceivable (physical) effects cannot be expected to occur immediately. Therefore, the relevance of the criterion of "immediacy" – although perfectly logical as such – could be minimised in hacking-cases showing a scope of sophistication that raises the political concern of a State in terms of *ius ad bellum*.

Directness

The criterion of "directness" describes the direct casual connection between the initial act and the resulting consequences of malicious cyber-activities. The explanatory text contrasts

27 N. Falliere, L.O. Murchu & E. Chien, *W32.Stuxnet Dossier* (Symantec Publication, Version 1.4, February 2011), at pp. 2 and 4, *available at* http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf (last visited 18 April 2012).

28 *Id*., at p. 7.

29 *See* D. Albright, P. Brannan & Ch. Walrond, *Stuxnet Malware and Natanz: Update of ISIS December 22, 2010 Report* (ISIS Report of 15 February 2011), at p. 3, *available at* http://isis-online.org/uploads/isis-reports/documents/stuxnet_update_15Feb2011.pdf (last visited 16 April 2012); Y. Katz, "Stuxnet may have destroyed 1,000 centrifuges at Natanz", in *The Jerusalem Post* online of 24 December 2010, available at http://www.jpost.com/Defense/Article.aspx?id=200843 (last visited 16 April 2012).

30 A denial of any physical damage by Iranian officials was reported by Reuters, "After Stuxnet: Iran says it's discovered 2nd cyber attack", in *The Jerusalem Post* online of 25 April 2011, *available at* http://www.jpost.com/IranianThreat/News/Article.aspx?id=217795 (last visited 16 April 2012).

the "directness" of the consequences of armed actions with the indirectness of e.g. economic coercion. This assessment is certainly true. However, the directness of the consequences of military actions can only apply to conventional kinetic operations – it is conceivable that the employment of BC-weapons in international relations, which will very likely always be considered "use of [armed] force", can show already a much weaker "directness" between their employment and the effects caused. The picture can change dramatically, if the range of remote weapon systems at the disposal of highly developed military forces, and especially the development of offensive military cyber capabilities, is considered.

However, the criterion of "directness" between the initial act and the resulting consequences seems problematic. The criterion determines, as Professor Schmitt rightly states, the conditions of attribution of certain perceptible consequences to a certain action in terms of causation (and maybe also a direct nexus?) between an action and the effects of that action. According to the understanding of the author of the present survey, the causation and direct nexus between an action and the effects of an action cannot be part of the assessment of the legal nature of the action as such. Therefore, the criterion of "directness" cannot be used for the classification of the nature of a malevolent action as being or not being "use of [armed] force". This opinion is certainly based on the different, rather dogmatic approach to the line of argumentation inherent to the civil law system.

Invasiveness

Subject to further discussion, it could be beneficial to clarify how the criterion of "invasiveness" shows relevance beside the criterion of "severity", the latter one describing the requirements of perceivable physical effects of malicious cyber-activities (not on the affected data only) in order to be likely to be categorised as "use of [armed] force". Especially, espionage by the means of the Internet or other information and communication technologies, i.e. illegal copying of data, would clearly be excluded from such a categorisation by applying the criterion of "severity".

Further, the same arguments and examples as demonstrated at the discussion of the criterion of "immediacy" are likely to apply to the criterion of "invasiveness", i.e. "invasiveness" of malicious cyber-activities could be imperceptible for a long period of time – for different reasons – and thus the relevance of the criterion could be minimised in practice.

Finally, it shall be mentioned that the criterion of "invasiveness" could show a certain potential for misuse, if the invasiveness of malicious cyber-activities were applied in the context of "national interest" when assessing malicious cyber-activities in the context of *ius ad bellum*. The prohibition of the "use of [armed] force" in international relations in the meaning of Article 2(4) of the UN Charter (and the right to self-defence, given in most severe cases of "use of [armed] force") does not protect national interests (see above).

Measurability

The criterion of "measurability" of effects of malicious cyber-activities, or rather their "appearance", is certainly an important one. It could be seen as complementing the criterion of "severity" of effects, although it might be beneficial for future discussions to further specify the relationship between these two criteria.

However, apparent effects of malicious cyber-activities will not always be measurable. For example, in the case of successful malicious cyber-activities against critical infrastructures of a State, apparent secondary, tertiary etc. effects, e.g. panic reactions within the population, disturbances of public order etc. (comparable to effects caused by e.g. a bombardment), will hardly be measurable. However, an aggressor who chooses a sophisticated way and modern means (i.e. malicious cyber-activities) for causing such effects of public disturbance, should not benefit from the fact that such effects are difficult to measure and, therefore, the classification of the actions as "use of [armed] force" could fail due to the requirement of the criterion of "measurability" of the effects. Therefore, indeed, the criterion should be used with caution.

Further, covering penetrations of computer systems and their negative effects by private companies, which own and operate the vast majority of critical infrastructure systems of a State (see above), could also minimise the relevance of the criterion in practice.

Presumptive legitimacy

Again, due to the rather dogmatic approach inherent to the civil law system, the criterion of "presumptive legitimacy" seems – from this perspective – problematic for several reasons:

First of all, "legitimacy" (describing an ethically justifiable act) is rather a term of political and ethical discourse; law deals with legality and illegality of actions. The judgement of (il) legality of actions inherently involves questions of (il)legitimate behaviour, but only in the understanding of the nature of law as reflecting commonly agreed norms of morality and ethics, and as far as the (international) law explicitly foresees an ethical assessment by an individual or a group of individuals (e.g. in regard to the determination of the term "excessive" or of the notion of "proportionality"). Further, assuming that legitimacy of an action indicates its legality, the criterion seems to contain a circular reasoning: The presumption of legitimacy cannot be part of the assessment of the legality. In other words, it cannot be decided whether a particular act is indeed legal under the *ius ad bellum* by the simultaneous assertion or indication of its legality at the same time. Moreover, it seems problematic to assume that legitimacy would have an impact on the assessment of the legality of an act (in our case: under the *ius ad bellum*). For example, in 1999 the military campaign in Kosovo, which was conducted without the consent of the State in question and without authorisation from the UN Security Council, and aimed to rescue a certain ethnic group likely to suffer ethnic cleansing, was determined by the "Independent International Commission on Kosovo" in its respective report as "illegal but legitimate".[31] This shows that (presumed or determined) legitimacy does not have an impact on the assessment of the legality. Last but not least, the "first sight" (or "jurisprudential intuition"?) of the legality of an action, indicated by the (subjective) perception of its legitimacy, cannot be part of a thorough legal assessment of a situation in question.

Even if the above considerations are ignored, the criterion shows potential for further discussion: The criterion of "presumptive legitimacy" shall help distinguish "use of [armed] force" from acts like propaganda, psychological warfare or espionage, which are not forbidden under the *ius ad bellum*. However, "psychological warfare", according to the understanding of the author of the present survey, can be conducted only as the first step of or in the course of an already ongoing military operation, i.e. after the threshold of *ius ad bellum* has been crossed. Therefore,

31 Independent International Commission on Kosovo, *Kosovo Report: Conflict, International Response, Lessons Learned* (Oxford, Oxford University Press, 2000), at p. 2.

the example of "psychological warfare" is not helpful in determining whether an activity would cross the abovementioned threshold. As for the examples of espionage and propaganda (the latter probably even if reaching the level of inciting insurgency against another State's government), the criterion of "severity" could already rule out those activities as constituting "use of [armed] force".

Responsibility

The criterion of [State] "responsibility" addresses an especially complex issue, which has already initiated many debates within the legal and political sciences and which will surely be of most importance in the future. A thorough discussion of the topic would certainly exceed the scope of this paper. Therefore, in the following, a few thoughts will be sketched, hopefully initiating future discussions in more depth.

Cyberspace enables (skill and knowledge-wise) super-empowered individuals and groups of individuals to cause the most severe physical effects through manipulations of computer systems that the functioning of highly developed post-industrial countries depends on. Due to the possibility to act anonymously in cyberspace and to masque and hide the data streams, it will probably always be a major challenge to attribute malicious cyber-activities to a State. The technical attribution as well as the legal attribution (in the meaning of obtaining tangible evidence in form of Internet protocols from all the servers, nodes and switches the data stream was passing on its way around the world) are very limited in cases of highly sophisticated cyber-activities. The political attribution has – in a way – more freedom of action, as it can work with factors like the assessment of the overall political situation and can apply e.g. the *cui bono* test. However, taking into account the supposed indirect and quiet use of "proxies", e.g. patriotic hackers (hacktivists), by certain States, invoking State responsibility for cyber-activities will very seldom meet the legal requirements as currently set by international jurisdiction and scholarly writings, i.e. the test of an "effective" or "overall" control of the State over the activities of the non-State actors.[32]

Considering the enormous difficulties in this context, it was proposed in diplomatic circles to introduce the principle of "due diligence" of States in regard to activities of non-State actors originating from the States' territories. Indeed, a principle of "due diligence" can be identified in public international law, as States do have the obligation not to let their own sovereign territory be used for activities causing damage to another State. Such a principle can be derived from the principles of sovereign equality of States and of good neighbourhood (see also Articles 2(1) and 1(2) of the UN Charter), and can be supported by several resolutions of the UN General Assembly (see e.g. Friendly Relations Resolution[33] and Definition of Aggression[34]). The obligations and rights deriving from such a "due diligence" principle are already expressed

32 The discussion of the control levels for actions of non-State actors in the context of State responsibility would certainly exceed the scope of the present paper. For further information *see* e.g. A. Cassese, "The Nicaragua and Tadic Tests Revisited in Light of the ICJ Judgment on Genocide in Bosnia", Vol. 18 *European Journal of International Law* 2007, pp. 649 *et seq*.

33 Declaration on Principles of International Law Concerning Friendly Relations and Co-operation among States in accordance with the Charter of the United Nations, A/RES/2625 (XXV) of 24 October 1970, Annex.

34 Definition of Aggression, A/RES/3314 (XXIX) of 14 December 1974, Annex.

in numerous international treaty provisions[35], in various States' declarations[36], and are endorsed by the jurisdiction of the ICJ[37] in regard to international environmental law. A "due diligence" in regard to cyberspace would surely involve the implementation of precautionary measures, including political, organisational, administrative, legal and technical measures in order to prevent the misuse of the possibilities that cyberspace offers for malicious activities by non-State actors harming other States. However, it is rather doubtful that violating the "due diligence" obligations would automatically lead to the responsibility of a State for all malicious cyber-activities originating in its territory without considering requirements that the current law of State responsibility sets.

It was also proposed during a conference to use the concept of "reverse of proof" as is known in many national legal systems. However, such a reverse of proof would establish a *prima facie* responsibility of a State for all malicious cyber-activities which seem to originate from the State's territory. This could lead to undesirable results. For example, despite the greatest efforts, the data stream between the worm *Stuxnet* and its creators could be traced the farthest to command and control servers located[38] in Denmark and Malaysia – States clearly not suspected to be responsible for the creation, implementation, control of and effects supposedly caused by *Stuxnet* in either legal or political terms.

The "safe haven" theory[39], developed in the context of Article 51 of the UN Charter in regard to terrorists acting from the territory of so-called "failed States" or States unwilling or unable to impede activities of non-State actors harmful to other States, would be a valuable thought also in regard to the State responsibility for malicious cyber-activities of non-State actors otherwise qualifying as "use of [armed] force" and enabling the victim State to legally conduct a range of possible retorsions and counter-measures. However, this approach would also not conform to the current law of State responsibility, thus further discussions within the international community will be necessary.

The question of whether individuals can trigger the right to self defence[40] could be relevant – in parallel – also in regard to the question of whether non-State actors could undertake activities otherwise judged as "use of [armed] force" and triggering the right of States to undertake retorsions and counter-measures. There are considerable pros and cons – their demonstration would, unfortunately, clearly exceed the scope of this paper.[41] Considering the power the

35 *See* an overview of treaties on international environment protection deposited with the UN at the UN Treaty Collection Website, *available at* http://treaties.un.org/Pages/Treaties.aspx?id=27&subid=A&lang=en (last visited 17 April 2012). It shall be mentioned that the overview does not contain (numerous) regional treaties, especially the ones on international regimes for the use of rivers, lakes and other territorial waters.

36 L. Gründling, "Environment, International Protection", in R. Bernhardt (ed.), *Encyclopedia of Public International Law* (Vol. II., 1995), p. 96 *et seq*., at p. 101.

37 *See* ICJ, *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, ICJ Rep. 1996, p. 226 *et seq*., at p. 241 *et seq*. para. 29; ICJ, *Gab_ikovo-Nagymaros Project (Hungary v. Slovakia)*, ICJ Rep. 1997, p. 7 *et seq*., at p. 41 para. 53. *See* also *Trail Smelter Case (United States, Canada)*, 16 April 1938 and 11 March 1941, in *United Nations, Reports of International Arbitral Awards*, (Vol. III, United Nations Publication, 2006), pp. 1905-1982, *available at* http://untreaty.un.org/cod/riaa/cases/vol_III/1905-1982.pdf (last visited 16 April 2012).

38 Falliere, Murchu & Chien, *supra* note 27, at p. 21.

39 For an overview on the major lines of argumentation see Schmitt, *supra* note 23, at p. 602 *et seq*.

40 *Id*., at pp. 600-602.

41 *See* e.g. Ziolkowski, *supra* note 26, at pp. 221-229, demonstrating the lines of interpretation of Article 51 of the UN Charter, of the respective international customary law, as well as of international jurisdiction, State practice and resolution practice of UN organs after the events of 9 September 2001.

Internet gives, especially to skilled and knowledgeable individuals, a respective discourse can very probably not be avoided in the future.

5. CONCLUDING REMARKS

"Use of [armed] force" is given in the case of malicious cyber-activities which (indirectly) cause (1) deaths or (2) physical injuries of living beings, (3) destruction of property or (4) medium to long-term disruption of critical infrastructure systems of a State, if the effects are equal to the physical destruction of the respective systems. When additionally showing a considerable scope and intensity of effects, such malevolent cyber-activities can be considered an "armed attack", triggering the right of a State to self-defence. The criteria thus stay – deliberately – vague.

Given the highly political nature of the question of whether "use of [armed] force" in international relations or an "armed attack" occurred and, subsequently, a State considers itself in the right to undertake either a range of unfriendly acts and counter-measures or self-defence measures, more meticulous criteria for such an assessment seem inappropriate. Even if States would develop internal guidance on such questions, it is likely that they would display a considerable grade of abstraction. Only such general criteria will leave enough room for political manoeuvring in a process of decision-making, which potentially can lead to political tensions, disturbance of international peace and security and – as *ultima ratio* – to the possible rigorous result of resorting to use of force.

Additionally, the effects-based approach to the question of whether particular malicious cyber-activities are to be considered "use of [armed] force" or an "armed attack" should lead to the conclusion that the criteria for a respective decision taken by a State will perfectly resemble those used to identify whether conventional military actions causing similar effects would be considered as comprising such situations. Therefore, there is no need for the development of special criteria for malicious cyber-activities going beyond those focusing on the effects (indirectly) caused.

The assessment of malicious or damaging activities, reaching the level of political concern, cannot make a difference according to the – rather conventional or rather modern – means used in order to cause the effects raising political concern. Therefore, only criteria referring to the effects caused should be considered appropriate.

The author of the present survey acknowledges that the proposed general criteria will not be useful in situations "[…] in which the necessity of self-defence is instant, overwhelming, leaving no choice of means, and no moment for deliberation."[42], i.e. in the situation of an immediate "armed attack" triggering the so-called preventive self-defence. This is based on the fact that – despite additional intelligence – the intended effect of malevolent cyber-activities will not be visible beforehand. Very likely, cases of (legal) preventive self-defence will stay theoretical. Moreover, judged from today's perspective, even in the case of discovery of malicious codes in e.g. governmental computer networks there still would be a "choice

42 Quoted in I. Brownlie, *International Law and the Use of Force by States* (Oxford, Clarendon Press, 1963), at p. 43.

of means" and a "moment for deliberation". Malware can be isolated, penetrated networks disconnected and IT-security measures directed at the targeted networks – instead of more drastic, including forceful, measures directed against the malevolent aggressor. At the end of the day, the prohibition of the use of force in international relations and the right to self-defence do not protect the interest in modernity and comfort of life, economic returns or other national interests as such. The threshold of endangering the (physical) security of a State is a high one and should not be diluted.

Finally, it shall be mentioned that in regard to the academic discussions, whether a certain category of a malicious cyber-activity can be considered "use of [armed] force" or "armed attack", the – otherwise very commendable – distinction between *lex lata* and *lex ferenda*, as stated by many scholars, might be not always be appropriate. A line of argumentation can only be presented *de lege ferenda* if it differs from the already existing law. The discussions, however, mostly examine how the already existing law applies to cyberspace. Indeed, the development of a common understanding of the interpretation of the *ius ad bellum* in regard to cyberspace is very much needed, in terms of both the scientific research and the use for political practice; academia and Professor Schmitt, especially, is to be congratulated for pioneering with benefit for both areas.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

The 'Use of Force' in Cyberspace: A Reply to Dr Ziolkowski

Michael N. Schmitt
International Law Department
United States Naval War College
Newport, U.S.A.
schmitt@aya.yale.edu

Abstract: This article responds to Dr Ziolkowsi's article Ius ad bellum in Cyberspace – Some Thoughts on the 'Schmitt-Criteria' for Use of Force. It discusses the distinction between the terms 'use of force' and 'armed attack' in an effort to situate the former as a legal term of art. The article concludes that as the meaning of use of force is uncertain, it is useful to identify those factors that States are likely to take into consider when faced with the need to characterize an action. Such factors may be legal in nature, but will also often reflect national security interests.

Keywords: *use of force, Article 2(4) UN Charter, Schmitt Criteria*

Over a decade ago, I had the occasion to consider the *jus ad bellum* implications of 'computer network attack' in an article published in the *Columbia Journal of Transnational Law*.[1] At the time, such operations were emerging as a new method of warfare, but international legal assessments thereof lagged far behind. Although the piece drew a degree of attention,[2] interest in cyber matters rapidly faded as transnational terrorism captured the international legal community's attention following the horrific '9/11' attacks.

The cyber operations mounted by hacktivists against Estonia in 2007, as well as employment of cyber operations during the international armed conflict between Georgia and Russia the next year, refocused attention on the subject. Since then, cyber issues have dominated discussions among international lawyers and international security specialists. In reaction to these and other cyber incidents, most notably the 2010 Stuxnet attack, States have formulated national cyber

1 Michael N. Schmitt, *Computer Network Attack and Use of Force in International Law: Thoughts on a Normative Framework*, 37 Columbia Journal of Transnational Law 885 (1999). The jus ad bellum is that aspect of international law that addresses when it is that States may lawfully resort to use force as an instrument of their national policy. It must be distinguished from the *jus in bello*, which concerns how hostilities may be conducted once an armed conflict is underway. The latter body of law is also labeled international humanitarian law.

2 The term was coined in Thomas C. Wingfield, The Law of Information Conflict: National Security Law in Cyberspace (2000).

strategies,[3] formed cyber military units,[4] and established international centres dedicated to examining cyber conflict.[5]

Of particular note in this regard is a Cooperative Cyber Defence Centre of Excellence (CCD COE) funded project to draft *The Tallinn Manual on the International Law of Cyber Warfare*. As director of the project, I have benefitted from the knowledge and insights of the group of 25 world-class international legal and technical experts who have been participating in the effort, which will conclude this summer. Among the topics they have explored is the legal notion of 'use of force'. In the process, the group submitted the so-called 'Schmitt Criteria' for the use of force that I had originally set forth in the *Columbia Journal* article to a rigorous peer review. I remain convinced that they are sound, at least when applied as I originally intended.[6]

My friend and colleague Dr Katharina Ziolkowski has graciously asked me to pen a reply to her impressive and insightful contribution to this volume in which she offers thoughts on my criteria. I am delighted to engage in this 'dialogue' with her and hopefully clarify my approach somewhat. It is comforting to know that our overall conclusions part ways only at the margins.

The question at hand is when does a cyber operation amount to a use of force in the *jus ad bellum* sense? The prohibition on the use of force is codified in Article 2(4) of the United Nations Charter. That article, which applies only to the actions by or attributable to States, provides:

> All Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations.

Article 2(4) must be juxtaposed to Article 51:

> Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security.

Self-defence as provided for in Article 51 constitutes one of the two universally recognised exceptions to Article 2(4)'s prohibition on the use of force by States (the other being an authorization or mandate to use force pursuant to Article 42)[7]; it is universally recognized as

3 See, e.g., Department of Defense, Strategy for Operating in Cyberspace (July 2011); White House, International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World (May 2011).

4 E.g., United States Cyber Command.

5 E.g., The Cooperative Cyber Defence Centre of Excellence, a NATO centre of excellence.

6 For my most recent discussion of the criteria, see Michael N. Schmitt, *Cyber Operations and the Jus ad Bellum Revisited*, 56 Villanova Law Review 569 (2011).

7 Although Article 2(4) refers to a prohibition on "use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations", it is clear that the prohibition extends to any use of force not authorized by the Charter. Originally, the draft Charter contained no reference to territorial integrity or political independence. Their subsequent inclusion was controversial; the "other manner" language was inserted to make clear that their inclusion was not meant to limit the article's reach.

reflective of customary international law.[8] The Charter's scheme is quite simple in the abstract – a State may *use force* when facing an armed *attack*.[9]

Note that the two articles employ different terminology – 'use of force' and 'armed attack'. The Charter's *travaux preparatoire* suggest that the difference was intentional. Negotiators at the 1945 San Francisco Conference, where the Charter was drafted and adopted, rejected the premises that 'force' was limited to 'armed' force and that actions qualifying as a use of force also necessarily qualified as an armed attack.[10]

In the *Nicaragua* case, the International Court of Justice addressed this carefully crafted distinction. It held that the terms embodied different legal thresholds. According to the Court, there are "measures which do not constitute an armed attack but may nevertheless involve a use of force". In other words, it is necessary to differentiate "the most grave forms of the use of force from other less grave forms".[11] The Court later reaffirmed the existence of this 'gap' in the *Oil Platforms* case.[12]

The distinction between the terms epitomizes the Charter's conceptual architecture. A horrendous conflagration initiated by States acting forcefully had just occurred, one resolved only through the collective action of other States. Accordingly, the Preamble identifies the key purpose of the United Nations as "sav[ing] succeeding generations from the scourge of war, which twice in our lifetime has brought untold sorrow to mankind". This purpose was to be accomplished by "unit[ing] our strength to maintain international peace and security".[13] In light of these aims, it made sense to set a low threshold for qualification as acts that seriously endangered international security (use of force), but a high one for qualification as acts that rendered unilateral forceful actions permissible (armed attack).[14] A Security Council empowered to authorize forceful actions by a United Nations military force would presumably police the gap between the two, that is, act in response to actions that amounted to a use of force, but not an armed attack.[15] Thus, while Article 2(4) establishes when a State has violated international law by using force, Article 51 permits the use of force as a remedy for States victimized by certain egregious uses of force known as armed attacks.

8 Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v. U.S.), 1986 I.C.J. 14, paras. 185-190 (June 27). Note that some scholars dispute whether the treaty and customary norms are identical. For the purpose of this article, any possible differences are irrelevant. The author takes the position that they are in fact identical.

9 States may also employ force in the face of an imminent armed attack under certain circumstances. Those circumstances do not bear on the points made in this article or Dr Ziolkowski's.

10 See U.N. GAOR Special Comm. on Friendly Relations, U.N. Doc. A/ AC.125/SR.114 (1970).

11 Nicaragua Case, *supra* note 8, paras. 191, 210.

12 Oil Platforms (Iran v. U.S.), 2003 I.C.J. 161, para. 51 (Nov. 6).

13 UN Charter, preamble.

14 The right of self-defence is interpreted by many States today as extending to acts conducted by non-State actors that meet the armed attack threshold. However, this premise is somewhat controversial. See Michael N. Schmitt, *Responding to Transnational Terrorism under the Jus ad Bellum: A Normative Framework*, in International Law and Armed Conflict: Exploring the Faultlines 157 (Michael N. Schmitt & Jelena Pejic eds., 2007).

15 UN Charter, arts. 43-49. The Charter also envisioned actions "by some of the members" that were to be authorized by the Council. This has proven to be the prevailing response. Article 51's right of individual or collective self-defence was but a fail-safe mechanism in the event the system could not respond quickly enough, a point illustrated by Article 51's authorization to act defensively only "until the Security Council has taken measures necessary to maintain international peace and security."

In light of this gap, it can be concluded that actions that do not qualify as an armed attack may nevertheless comprise a use of force. But what do the two terms mean? I have contended elsewhere that an armed attack is an action with consequences that involve death or injury of individuals or damage to objects.[16] Unfortunately, the meaning of the term use of force is more problematic. The Charter's text provides no guidance beyond structurally indicating that any armed attack is equally a use of force. Charter *travaux preparatoire* and subsequent events are of some assistance in that they demonstrate that the notion generally excludes economic or political coercion.[17] The *Nicaragua* case also provided examples of actions that qualify uses of force (arming and training guerrillas fighting against another State), and that do not (merely funding them).

What seems clear is that while all coercive actions are not uses of force, a use of force need not be armed (or necessarily even directly related to armed actions). Beyond this broad deduction, the criteria for qualification as a use of force remain abstruse. This uncertainty led to my proposal of the 'Schmitt Criteria'.

The criteria – severity, immediacy, directness, invasiveness, measurability, presumptive legitimacy, and responsibility – are replicated in the article by Dr Ziolkowski and need not be described in detail here. However, before turning to my reflections on her comments, it is essential to grasp how the criteria were intended to be used … and how they were not.

Despite the absence of a consensus understanding of the term use of force in either judicial pronouncements or State practice, States will sometimes be compelled to assess cyber operations against the prohibition set forth in Article 2(4) and contained in customary international law. At times, they will have to do so with respect to cyber operations conducted against them in order to decide whether to characterize the initiating State's actions as a violation of the norm. Sometimes, they will need to resolve whether other States will characterize cyber operations they are contemplating as a use of force. And in still other cases, they will have to assess cyber operations targeting other States. In light of the definitional lacuna described above, perhaps the best States can do is to engage in educated conjecture as to how the international community is likely to view the cyber operations in question as a matter of law.

The criteria were meant to assist in that effort. What is often misunderstood is that they are not legal criteria against which to perform such evaluations. For instance, the criteria do not have the legal status that the necessity, proportionality, and imminency/immediacy requirements associated with taking action in self-defence enjoy. Rather, they are merely factors that can be expected to influence States when making use of force appraisals. After all, what matters in international relations is not whether the actions in question are lawful in the abstract, but instead whether the international community considers them as such. Lest this assertion seem extreme, recall that customary international law is formed through the confluence of State

16 See my other article in this volume, *'Attack' as a Term of Art in International Law: The Cyber Operations Context.*

17 U.N. Doc. 2, G/7(e)(4), 3 U.N.C.I.O. Doc. 251, 253-54 (1945); Declaration on Principles of International Law Concerning Friendly Relations and Co-operation Among States in Accordance with the Charter of the United Nations, G.A. Res. 2625 (XXV), U.N. Doc. A/RES/8082 (Oct. 24, 1970); U.N. GAOR Special Comm. on Friendly Relations, U.N. Doc. A/AC.125/ SR.114 (1970); Report of the Special Committee on Principles of International Law Concerning Friendly Relations and Co-operation Among States, U.N. Doc. A/7619 (1969); Schmitt, *supra* note 6, at 574.

practice and *opinio juris*.[18] Consequently, an action that may have been unlawful in the past, but which is not viewed as such in the present, contributes to the eventual emergence of new customary norms.

Both the absence of meaningful State practice as to cyber operations (and the reaction thereto) and the definitional vagueness regarding the prohibition of the use of force signal that the law regarding the use of force in the cyber context is ripe for this evolutionary process. Therefore, for the immediate future, we can expect a period of relative flexibility in the application to cyber operations of the prohibition. As State practice accompanied by expressions of *opinio juris* develops, the law will slowly crystallize. Once this happens, prognostic criteria such as those I have proffered will be replaced by tangible legal requirements. In the meantime, States will continue to be influenced in their decisional process by factors like those I have suggested. They are, of course, non-exclusive, and their relative influence in matters of international security, including legal assessments as to State behaviour, is always contextual.

It should be evident that I am more cautious than Dr Ziolkowsk with regard to characterizing the force contemplated in the prohibition as 'armed'. Nevertheless, despite terminological divergence, we arrive a roughly the same conclusion as to force which causes physical harm to individuals or damages objects. They are uses of force as a matter of law, since I would equally assert that they meet the higher threshold of armed attack.

Interestingly, Dr Ziolkowski also characterizes cyber operations resulting in "massive, medium to long-term disruption of critical infrastructure systems of a State" as uses of force, at least to the extent that their effects are equal to the physical destruction of the system. I am somewhat less confidant than she is in this respect, in part because I am unsure that States will readily equate non-kinetic effects with kinetic ones. In my view, only State practice can establish such a 'bright line' norm. While agreeing that States may well characterize such actions as uses of force based on the criteria I have set forth (and other contextual factors), I am uncomfortable offering the standard as *lex lata* at this time. It may mature into either customary law or a customary interpretation of Article 2(4) over time, but the uncertainty attendant to cyber operations leaves her proposed standard currently fixed in the realm of *lex ferenda*.

As to Dr Ziolkowski's assessment of the criteria, I fear that she attributes rather more normative significance to them than I do. As noted, they are predictive tools, not normative standards. In this regard, I might suggest that the differences between our approaches derive less from our differing civil and common law backgrounds than from the different perspectives we have towards international law. Whereas she adopts an approach based primarily in positivism, mine reveals the influence of the policy-oriented New Haven School.[19] For me, law, contextually understood, often reflects policy choices that are shaped to achieve particular values. This explains my readiness to identify influences on legal assessments that are not strictly legal in nature. Thus, while both our approaches are consequence-based, she pays greater attention

18 Statute of the International Court of Justice, art. 38(1)(b). See also North Sea Continental Shelf (Federal Republic of Germany v. Denmark; Federal Republic of Germany v. the Netherlands),1969 I.C.J. 3, para. 77 (Feb. 20).

19 The intellectual fathers of the New Haven School were Yale Professors Myres McDougal and Harold Lasswell. It was later championed by such scholars as Michael Reisman. The first piece setting forth the approach was Harold D. Lasswell and Myres S. McDougal, *Legal Education and Public Policy: Professional Training in the Public Interest*, 52 Yale Law Journal 203 (1943).

to the nature of the consequences caused by cyber operations, whereas I tend to focus on the policy perspectives States are likely to have vis-à-vis those consequences.

Our differing normative vectors are revealed in Dr Ziolkowski's comments on what we agree is the most important criterion, severity. She rejects my assertion that the more cyber operations impinge on critical national interests, the more likely States are to characterize them as uses of force. For her, international law protects the physical security of a State and its inhabitants, not the State's national interests. By my policy-oriented approach, however, national interest is the most determinative factor. The very reason States accede to (or reject) international legal regimes is to protect those interests and the various values they reflect.

Security interests may dominate in *jus ad bellum* matters, but they are not exclusive. For instance, Article 2 of the United Nations Charter expressly states that the instrument is intended to foster the purposes set forth in Article 1. Beyond the maintenance of international peace and security, these purposes include the development of "friendly relations among nations based on respect for the principle of equal rights and self-determination of peoples" and the achievement of "international cooperation in solving international problems of an economic, social, cultural, or humanitarian character, and in promoting and encouraging respect for human rights and for fundamental freedoms for all without distinction as to race, sex, language, or religion".[20] By the approach I have espoused, it is less the nature of the national interest than its intensity that matters.

An analogous thread runs through the immediacy criterion. Immediacy influences decision-makers because it heightens the need for a victim State to characterize the situation quickly lest the negative consequences of a cyber operation manifest themselves before the State can muster the domestic and international support necessary to validate any responsive action it might take. This will force the hand of other States, which in the majority of cases will be predisposed to a characterization benefitting the victim State. After all, because the cyber operation is likely to be unlawful irrespective of whether it amounts to an unlawful use of force, doubt is likely to be resolved in favour of the victim. Operations that only generate effects over the medium or long term, on the other hand, afford all parties a greater opportunity to rule out the possibility that the originator State has engaged in a use of force.

The other criteria will similarly influence assessments in ways that exceed their technical legal valence. For example, the more direct the casual connection between a cyber operation and its impact on national interests, the more comfortable States will be in describing the operation as a use of force. The law aside, portraying an originator State's actions as a violation of the *jus ad bellum* is a politically charged step, one that always presents political risks. Directness can serve to mitigate such risks by shifting the onus of responsibility for disrupting peace and security to the originator State. The same holds true with regard to invasiveness. The more invasive a cyber operation, the less politically risky the act of asserting that the originator State has used force in contravention of international law. In the case of cyber operations that are particularly direct and/or invasive, the victim State will also feel more aggrieved, thereby making it readier to style the operations as a use of force, a characterization with which other States are likely to sympathize.

20 UN Charter, arts. 1(2) & (3).

Measurability and the lack of presumptive legitimacy will also make use of force characterizations easier to defend before both domestic audiences and the international community because the victim State and those States that support its characterization can offer hard facts to justify (as a matter of fact and law) their determination without having to rebut any presumption of legitimacy. Finally, although the legal issue is qualification as a use of force rather than State responsibility for the use of force, the victim State and its supporters will be more comfortable alleging that the originator State has engaged in a use of force when the latter is clearly responsible pursuant to the principles of State responsibility.

The point is that the factors set forth will, legal considerations aside, influence States when making determinations regarding an area of unsettled law. After all, States understandably tend to resolve uncertainty in favour of that position that best advances their interests. A State victimized by a cyber operation will usually deem it in its interest to assert that the delict has been severe, whether to engender sympathy or to generate support for any responsive measures it might wish to take. Uninvolved States are in a somewhat different position. In particular, a State that anticipates conducting similar cyber operations itself has an incentive to characterize analogous actions by other States as falling short of a use of force. Nevertheless, as a general rule, uninvolved States are more likely to accept the characterization of the victim State as reasonable the more the criteria set forth are met. This is especially so when they see themselves as potential victims of cyber operations.

Reduced to basics, the 'Schmitt criteria' represent an acknowledgement of the ambiguity resident in the use of force norm. Given the ambiguity, the decisional latitude of States is wide. They will inevitably leverage this decisional flexibility by adopting legal positions that optimize their national interests. The criteria are what I believe to be some of the key extra-legal influences on that complex process. Accordingly, they are meant to be predictive, not prescriptive.

I am grateful to Dr Ziolkowski for opening the dialogue about the 'Schmitt Criteria' to a wider audience, especially those concerned with the technical and policy aspects of uses of cyber force. She is to be applauded for offering a sophisticated assessment of them and I am sincerely appreciative to her for the opportunity to clarify my thoughts.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

A Methodology for Cyber Operations Targeting and Control of Collateral Damage in the Context of Lawful Armed Conflict

Robert Fanelli
United States Cyber Command
Fort Meade, Maryland, USA

Gregory Conti
United States Military Academy
West Point, New York, USA

Abstract: Throughout history, the law of warfare has evolved to protect non-combatants and limit collateral damage. The same legal and ethical constraints apply to the conduct of cyber warfare, where it is similarly desirable to limit the effects of offensive actions to specific locations and groups. However, conventional wisdom suggests that this is extremely difficult, if not impossible to accomplish in the cyber domain. In this paper, we argue to the contrary. It is possible to constrain the effects of cyber actions to specifically desired, legitimate targets while significantly limiting collateral damage and injury to non-combatants. To this end we present a generalized methodology for analysis of the targeting and effects of cyber operations with respect to principles of lawful conduct in armed conflict. This methodology includes a framework of effects categories, target attributes and control measures to direct and constrain cyber operations. It also includes a process for evaluating these effects and controls against the principles for lawful conduct in armed conflict. We illustrate the methodology in action by applying it to W32.Stuxnet, software widely considered to be a cyber weapon. Our results indicate that it is entirely possible to analyze complex cyber war problems, identify legally authorized courses of action, and focus effects on desired targets while greatly minimizing collateral damage.

Keywords: *cyber operations, targeting, collateral damage, law of armed conflict*

1. INTRODUCTION

While unfortunate, armed conflict has existed since the dawn of man. Over time, customs, agreements and laws have evolved to define what actions are permissible and prohibited in armed conflict. For example, among other requirements, humanitarian law imposes a duty on combatants to avoid injury to non-combatants and to limit collateral damage [1]. In general,

The views expressed in this article are those of the authors and do not reflect the official policy or position of the US Cyber Command, United States Military Academy, Department of the Army, Department of Defense, or United States Government. All information and sources for this paper were drawn from unclassified material.

standards for behavior in armed conflict on land, at sea or in the air are well-understood, having evolved over many years.

A similar understanding for warfare in the cyber domain does not yet exist. Much work has examined the legal aspects of operations in the cyber domain, attempting to reconcile such operations with existing notions of what constitutes armed conflict or an act of war [2,3,4]. However, the literature is mostly silent on how we may actually execute cyber operations in a manner that complies with accepted standards for conduct in armed conflict in particular. Some believe that constraining the effects of cyber operations is technically infeasible given the complexity and interconnectedness of information systems and networks, making all such operations illegal [5]. We argue that it is indeed possible to comprehensively study the operational factors and conduct cyber operations within legal and ethical constraints while achieving legitimate military objectives.

In this paper we make several contributions. We introduce a methodology to categorize the effects of cyber operations. We also present a framework of target attributes and control measures to direct and constrain cyber operations. Finally, we present a general methodology for evaluating these effects and controls against the principles for lawful conduct in armed conflict.

This paper is organized as follows. Section 2 places our research in the field of related work. Section 3 presents our generalized methodology. Section 4 examines application of the methodology to a cyber operation. Section 5 presents our conclusions and promising directions for future work.

2. BACKGROUND AND RELATED WORK

There is a great deal of interest in the opportunities and challenges of conducting military operations in cyberspace. A number of definitions exist in the literature for the term 'cyberspace.' For this work, the definition proposed by Daniel Keuhl is suitable: "…cyberspace is a global domain within the information environment whose distinctive and unique character is framed by the use of electronics and the electromagnetic spectrum to create, store, modify, exchange, and exploit information via interdependent and interconnected networks using information-communication technologies." [6]

A commonly held view of cyberspace is one limited to computer systems connected by the TCP/IP-based Internet. However, Keuhl's definition of cyberspace includes a larger set of systems, protocols, architectures and functions including, but not limited to, those found on the Internet. Thus it is important that our discussion is sufficiently general, and our results are sufficiently flexible, to address the full range of information systems, networks and transmission media in cyberspace. Still, the framework must retain sufficient specificity to inform actual offensive action in the domain.

The United States Department of Defense and others have only recently recognized cyberspace as a separate domain of armed conflict besides land, sea, air and space [7]. Despite being a

distinct domain with unique qualities, the cyber domain requires standards for lawful behavior in armed conflict just as do the other domains [1]. Despite progress, we have not yet fully determined how the customs and body of law that define acceptable behavior in armed conflict will apply to the cyber domain.

Several authors have attempted to relate concepts of strategic warfare and deterrence to the cyber domain. [8,9,10] However, these works do not address nature of offensive action in the domain at the operational and technical levels.

A number of authors have also discussed legal aspects of cyber operations. Much that has emerged from this discussion, such as the 'Schmitt Criteria' [4], is concerned with the role of cyber operations in terms of *jus ad bellum*, or determining when resorting to war is justified and what constitutes an act of war.

One point of debate is whether or not cyber operations can, in fact, constitute armed conflict. Sklerov proposes "an effects-based approach, sometimes called a consequence-based approach, in which the attack's similarity to a kinetic attack is irrelevant and the focus shifts to the overall effect that the cyber attack has on a victim state." [2] Cyber operations do in fact amount to armed conflict when their effects are consistent with those of more established, kinetic forms of armed conflict, highlighting the need to pay particular attention to the potential effects of cyber operations.

For this work, we set aside the concerns of *jus ad bellum* and focus instead on *jus in bello*, the rules for lawful conduct of armed conflict *after* the decision to resort to military action is made. *Jus in bello* imposes duties to use restraint in the application of force, minimize suffering, and distinguish between legitimate military targets and non-combatants when conducting attacks. Further, combatants have a duty to control the collateral damage that may result from military operations. [1,3]

Sklerov identifies four principles of *jus in bello*:

1. Distinction: combatants have a duty to ensure attacks are directed at legitimate military objectives and to minimize collateral damage.
2. Necessity: the application of force must be limited to only the amount necessary to accomplish a valid military objective.
3. Humanity: weapons designed to cause unnecessary suffering are prohibited.
4. Proportionality: limits the use of force to situations in which the expected military advantage outweighs the expected collateral damage to civilians and their property. This does not require avoiding all collateral damage; rather, such damage must not be out of proportion with military necessity [2].

In addition to these principles, Schmitt cites a principle of *discrimination* [1]. This prohibits the use of 'indiscriminant' weapons or tactics, those incapable of avoiding damage to non-combatants.

These well-established principles dictate that one must be able to precisely target and control the effects of the weapons and techniques employed in armed conflict. Beyond having sufficient control, one must also ensure operations target valid military objectives in accordance with the *jus in bello* principles. "Those who plan or decide on attack have an affirmative duty to 'do everything feasible' to verify that intended targets are legitimate." [1]

For this work, we shall focus on the principles of discrimination, distinction and proportionality. Methods for ensuring cyber operations adhere to these three principles differ most from those for kinetic operations, posing the most significant challenges. The principles of necessity and humanity are similar in both kinetic and cyber operations. Compliance will follow from meeting the challenges posed by the other three principles.

The methods to discriminate between combatant and non-combatant and to reduce collateral damage in the kinetic domains of land, sea, air and space are relatively well understood. The effects of actions in the kinetic domains tend to be well localized in physical space. Similarly, physical science and modeling provide accurate predictions about the duration and spread of such effects. Admittedly, achieving these goals in practice is not always easy, mostly due to 'fog of war' and limited intelligence about the true nature of a target.

In the cyber domain, measures of location, distance and time may be less effective for ensuring compliance with the principles of *jus in bello* than they are in the physical domains. We also have far less history and experience dealing with the questions of how to target and constrain effects in the cyber domain. However, the requirement to conduct cyber operations in a manner consistent with *jus in bello* remains. Thus there is the need for a methodology, such as that presented here, to analyze cyber operations effects, targeting and control measures in terms of the lawful application of force in armed conflict.

3. A METHODOLOGY FOR CYBER OPERATIONS TARGETING AND CONTROL

'Cyber weapons,' and those wielding them, must be capable of operating in accordance with the principles of *jus in bello*. This entails the capability to direct effects at valid military targets using controlled amounts of force and to minimize collateral damage. Organizations conducting cyber operations require sufficient intelligence capabilities for accurate targeting plus agile and robust command processes to control and to accurately assess their effects. With respect to tools, these requirements differentiate cyber weapons from the more general category of malicious software, or malware. Malware is frequently indiscriminant and poorly controlled, seeking to spread and cause effects as widely as possible with little regard for the nature of the victims. The methodology presented here seeks to provide a framework in which those from the technical, legal and policy making disciplines can achieve consensus on lawful conduct for specific cyber operations and weapons.

A. Cyber Operations Effects

The potential severity and scope of a cyber weapon or operation's effects dictate the degree of control needed to act in accordance with *jus in bello* principles. Operations and weapons

capable of causing more severe damage, or with consequences more widespread in space and time, call for greater precision in targeting and control of effects. Thus we must have means to categorize the severity and persistence of effects. We define three categories of severity for effects:

Primary effects have the potential of directly affecting physical assets and human lives. This would include manipulating control systems to cause the malfunction of machinery, power outages, explosions, flooding, vehicle accidents or other physical destruction. It also includes rendering information systems and other electronics inoperative at the hardware and firmware level.

Secondary effects degrade or disrupt physical assets as a second-order consequence of effects in the cyber domain. Although a secondary effect does not have the immediate potential for direct physical destruction, it is still expected to affect physical assets. The disruption of information systems and networks in the cyber domain can affect physical assets reliant on them for control, monitoring and communications. Examples would include spoofing air defense systems, disabling telecommunication systems, incapacitating control systems for transportation or logistical networks, corrupting databases and manipulation of financial systems.

Indirect effects remain within the cyber domain, having only an informational impact. Attacks with indirect effects primarily impact human cognition and would be expected to affect the physical domain only through humans acting on the information perceived. Cyber operations having indirect effects would include military deception operations, delivering targeted messages to a populace and blocking or altering an adversary's messages.

A cyber weapon or operation may have the potential for causing multiple effects, possibly causing differing combinations of primary, secondary and indirect effects on different targets. We must consider each likely combination of target and effect for compliance with *jus in bello* principles.

We also define three degrees of persistence for effects:

Permanent. This level of persistence includes effects that require replacing hardware or extensive, time-consuming repairs. It also includes destruction of primary data and backups such that timely restoration is infeasible. Such effects would include disabling hardware through destruction of firmware, destruction of electronics through overloads, physical destruction of infrastructure or other property and loss of life.

Temporary. Temporary effects also persist after the operation ends; however, unlike permanent effects, recovery here entails actions of lesser cost in resources and time. Such procedures would fall within the scope of typical disaster recovery plans [11]. Examples include restarting disrupted telecommunications or electrical infrastructure, reloading operating systems and restoring data from backup media.

Transient. Transient effects abate quickly after the attack ends, with little effort on the part of the targeted entity. At most, recovery might include resetting or rebooting equipment. For

example, denial-of-service and traffic redirection attacks typically generate transient effects.

B. Target Attributes and Control Features

Cyber weapons and operations must have sufficient precision to ensure effects reach the intended targets while avoiding noncombatants. We require a flexible means to describe targets for the purpose of directing and constraining effects. We define three *target attributes* for cyber operations. Taken together, these attributes allow us to answer the questions: "Where is it?", "What is it?" and "Whose is it?" for a given target.

Geography. This attribute addresses the physical location of the target. This may be pertinent for two reasons. First, physical location within a given region, such as a national border, may define what is and is not a legitimate party to a conflict. Second, physical location may contribute to establishing a positive identification of the target, especially in ensuring it is not an entity with protected status and thus off-limits to attack. A geographic attribute may be as specific as a building, military installation or industrial facility or as broad as a nation or a military theatre of operations. The dynamic nature of networks and mobile devices may, in some circumstances, make determination of physical location difficult.

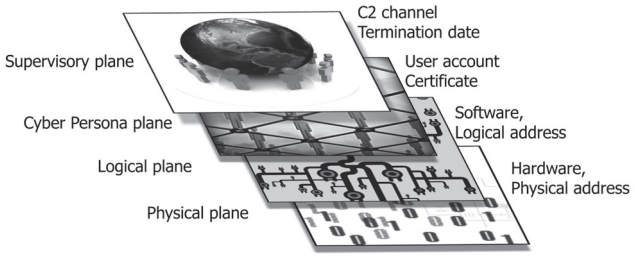
Function. This attribute addresses the purpose or current activity of a target. Identifying function provides a useful means to differentiate a legitimate target from other, nearly identical systems. For example, an industrial controller of a given type might be used for a humanitarian purpose in one location and a purely military role, such as producing munitions, in another. Further, combatants and protected entities could be intermingled on a shared network, in a cloud infrastructure or, through virtualization, on a single host. The information systems and networks that comprise cyberspace, by their flexible nature, may be the epitome of dual-use objects, potentially serving both civilian and military purposes [1]. A change of software or configuration could allow an information system to rapidly change function between civilian and military purposes. Thus, identification of a device's function may facilitate distinguishing intended, legitimate targets from others using similar hardware or software.

Persona. This attribute addresses the ownership and users of target. A persona attribute could relate to a person, business, government or other group. The personas involved with a given system or network may assist in identifying the intended target and separating legitimate targets from others. A persona may also be the primary descriptor for a target, indicating a person or group to be engaged wherever found in cyberspace, with less emphasis on geography and function.

After identifying the target in terms of geography, function and persona attributes, we must determine the specific information and technical features necessary for effective targeting and control of effects. The objective is to derive a set of *control features* sufficient to direct the effects to the intended target while avoiding disproportionate collateral damage or other unlawful consequences. Control features are specific values that a cyber weapon or operation can use to determine if effects should be delivered to a potential target device. The control features must be sufficiently general to encompass not only the TCP/IP based features frequently discussed in 'computer network attack' [12] but also the larger set of features available in the

array of devices and networks implied in the definition of the cyber domain. We divide the control features into four *planes*.

FIGURE 1. EXAMPLE CONTROL FEATURES BY PLANE



Physical Plane. The physical plane includes features of a device's hardware, its operating characteristics and its physical environment. Information about a device's hardware may identify its general type, its manufacturer or specific model, or possibly device unique identification by distinctive values such as embedded serial numbers.

Some devices, such as 'smart phones,' may provide direct information about their geographic locations through Global Positioning System or mobile network location services. Other physical features, such as clock settings, power sources and keyboard layouts may permit inferences to be drawn about the location of a device.

Physical features may provide information about the function of a device. Patterns of utilization, workloads, transmit and receive frequencies, function-specific firmware or environmental conditions may indicate a device's function and help differentiate it from similar devices. Similarly, the physical characteristics of devices attached to industrial control systems may provide information about the function or even the specific identity of the system.

Physical features that uniquely identify a device provide the potential for extremely precise targeting and control. Such information could tie the device to its owner or other persona and may also facilitate determining its function. Examples include serial numbers of hardware installed in computer systems or the International Mobile Equipment Identifier (IMEI) of a mobile device. Similarly, network address information associated with hardware on a persistent, if not permanent, basis can also provide device identification. Such features include medium access control (MAC) addresses for network interfaces, Mobile Identification Numbers (MIN) and International Mobile Subscriber Identity (IMSI) values on SIM cards.

Logical Plane. The logical plane includes features of the software on a device plus the configuration and state of that software. Primary examples are logical network addresses, such as Internet Protocol (IP) addresses. Although an IP address itself does not contain location information, the nature of IP networks and knowledge of address range assignments often make it possible to determine geographic location or ownership [13]. Other configuration items, such as time zone or language settings, may also help to infer a device's location.

Logical control features may also facilitate identifying the function of a device. The operating system and application software present, the configuration and state of the software and the data files, log files and other content stored on a device may differentiate devices having common hardware but serving different functions. Similarly, the programming of an industrial control system can provide useful information about the function of the system and possibly indicate its location and ownership.

Cyber Persona Plane. Cyber personas are identities in the cyber domain. These features are useful in determining the ownership, affiliation and users of a device. Physical personas and cyber personas often exist in one-to-many or many-to-many relationships. A person may have multiple cyber personas while a single cyber persona may in fact represent multiple, loosely related persons. An example of the latter case is the group 'Anonymous' [14].

The primary cyber persona control features are the user accounts on a device. These may include accounts for local and remote systems plus network services such as electronic mail. Cyber persona control features also include digital certificates, software license registration entries and stored biometric data. It may even be possible to capture images and audio from embedded cameras and microphones to definitively identify the user of a device.

Supervisory Plane. The supervisory plane contains the command and control features available to start, stop and redirect a cyber weapon or operation. This includes features related to human-in-the-loop command and control of targeting and effects during the operation. It also includes predefined trigger events for starting, stopping or changing some aspect of an operation and controls on the ability of cyber weapons to propagate autonomously.

This plane also includes temporal specifications for the timing and duration of effects. These may be specific start and stop times for operations or a duration limit for effects initiated in response to a trigger event.

C. Methodology for Enumeration and Analysis

Using the framework of effects, target attributes and control features presented above, we may now determine if a given cyber weapon or operation complies with *jus in bello* principles. This methodology involves considering the probable consequences of the operation against its precision in targeting and control.

First, we enumerate the likely primary, secondary and indirect effects of the cyber weapon or operation, along with the degree of persistence for each, on an *effects tableau*. In particular, any significant potential to cause death, bodily injury or destruction of property must be examined. Table I depicts an effects tableau with example entries.

After we have enumerated the likely effects of the cyber weapon or operation, we must examine its control features. This evaluation facilitates the military commander's determination if a planned operation complies with *jus in bello* principles. Alternatively, such analysis could be used during development to identify control features needed to ensure the cyber weapon produced is sufficiently precise to avoid unintended targets and limit collateral damage. We enumerate the control features of the cyber weapon or operation on a *targeting and control*

tableau, listing each control feature by its plane and the targeting attribute to which it pertains. Table II depicts a targeting and control tableau with example entries.

We now analyze the enumerated effects and control features. The goal is to determine if the cyber weapon or operation has sufficient control in terms of Geography, Function and Persona so that its effects are in accordance with the *jus in bello* principles of Discrimination, Distinction, and Proportionality. Considerations of proportionality in a cyber operation should compare with those for kinetic operations. If we would reject some possible collateral damage from a bomb or other kinetic effect, we should reject the same possibility if posed by the cyber operation. Conversely, risks of collateral damage found acceptable for kinetic operations should be similarly acceptable from cyber operations.

If we find the operation complies with the *jus in bello* principles for all its anticipated effects, the operation may lawfully proceed. On the other hand, if we identify noncompliance for one or more effects, it may be possible to modify the control measures to bring the operation into full compliance. Alternatively, it may be necessary to defer operations against a given target until tools and techniques offering sufficient control for their effects are developed or procured. Finally, we may conclude that a given combination of cyber weapon or operation and target do not comply with *jus in bello* principles and that we should consider other alternatives for achieving the military objective.

4. APPLICATION OF THE METHODOLOGY

To further illustrate our methodology, we apply it to W32.Stuxnet, software widely considered to be a cyber weapon. Stuxnet appears to be the best publicly-disclosed example of a potential cyber weapon, with detailed technical analysis readily available [15]. Multiple authors allege that Stuxnet was part of a cyber operation conducted by a state-level actor with the objective of sabotaging Iran's uranium enrichment program. [16-18] Although uncertainty remains about the origin and purpose of Stuxnet, we will assume here that the cyber attack explanation is correct.

We leave to others the question of the lawfulness this operation under *jus ad bellum*. Questions that remain are then: did an attack using Stuxnet constitute lawful armed conflict? Did this cyber weapon include sufficient precision and control of its effects to comply with the *jus in bello* principles of discrimination, distinction and proportionality?

A. Enumeration of Effects

First, we enumerate the likely effects of the operation. Stuxnet exploited multiple vulnerabilities in Windows operating systems to propagate, specifically targeting systems running the Siemens WinCC and SIMATIC Step 7 industrial control system (ICS) software used to manage programmable logic controller (PLC) devices. [15] Stuxnet's primary effect was to alter the operation of certain models of frequency controller, causing them to run the attached device at a very high speed and suddenly bring it to a near stop. This would be likely to damage or destroy devices such as high-speed centrifuges. Altering the intended operation of the frequency

controllers would also have the potential secondary effect of degrading the industrial process controlled. Such manipulation would significantly reduce the yield for a sensitive process such as uranium enrichment. [19]

As secondary effects, Stuxnet replaced or altered components of the WinCC and SIMATIC Step 7 software. Although Stuxnet implanted itself on Windows systems, it had no significant effects on those systems unrelated to gaining access to the target PLCs. Stuxnet also altered frequency converter activity data returned to management systems, ostensibly to mask indications of the primary effects. Table I depicts the effects tableau for Stuxnet.

We infer indirect effects for this operation since these are not coded in Stuxnet. Successful sabotage of the production process could result in a loss of confidence in the reliability of hardware, software and management processes, at least temporarily. A more permanent indirect effect is the possible loss of skilled personnel blamed for production losses or failing to prevent the attack.

TABLE I. EFFECTS TABLEAU FOR STUXNET

	Persistence		
Effect Class	Permanent	Temporary	Transient
Primary	Damage or destroy certain high-speed industrial devices		Alter operation of certain frequency controllers
Secondary	Sabotage industrial process dependent upon precise frequency controller operation	Affect Windows system integrity. Alter components of WinCC and Step 7	Deceive management systems by altering feedback from frequency converters
Indirect	Dismissal or criminal sanctions against management and staff	Loss of confidence in hardware, software or procedures	

B. Enumeration of Target Attributes and Control Features

We must now consider the target attributes and enumerate Stuxnet's control features. As stated above, we accept the hypothesis that the target of Stuxnet was Iranian uranium fuel enrichment facilities. More specifically, the target devices were the industrial control systems and IR-1 centrifuges employed in the uranium enrichment process [20]. What, then, were the attributes of this target?

Geography: the target was known to be located in Iran. Forensic analysis indicated that the initial infections occurred in five Iranian networks, probably from direct connection of portable storage devices [15].

Function: The target devices were industrial control systems carrying out the uranium enrichment process. This required the presence of distinctive controller hardware configurations and specific software to manage and monitor the process. Additionally, the process would

involve behavior, such as high rotational speeds for extended periods of time, differentiating it from more mundane functions.

Persona: the targets were owned and operated by Iranian government entities.

Stuxnet contained multiple features apparently designed to limit its effects to the intended targets. It is likely this was done as much for stealth as to control collateral damage; nonetheless, the controls were included. The most significant control features are related to the target's function and fall within the physical and logical planes. This is understandable since the function of the target in this case is significant and provides more specificity than geography or persona attributes. Stuxnet checks for specific ICS software, hardware and mode of operation before delivering its effects. Stuxnet also includes control features on the supervisory plane that provide some limits on propagation and basic command and control capability [15]. Table II depicts the targeting and control tableau.

TABLE II. TARGETING AND CONTROL TABLEAU FOR STUXNET

	Target Attribute		
Plane	Geography	Function	Persona
Physical	Initial launch via external storage devices connected to five Iranian networks	• Hardware: check for a Siemens PLC, type 6ES7-315-2, using a Profibus communications processor module CP 342-5 • Configuration: The PLC must be connected to at least 33 frequency controllers manufactured by either Fararo Paya (Iran) or Vacon (Finland)	N/A
Logical	N/A	• Software selectivity: Infect only Simatic manager (s7tgtopx.exe) and WinCC project manager (CCProjectMgr.exe) on Win32 • ICS operation: trigger primary effects only if specific operating pattern is observed. (Must operate at 807 Hz to 1210 Hz for 12.8 days, initially)	N/A
Cyber persona	N/A	N/A	N/A
Supervisory	N/A	• Copy limit: after three copies from an external storage device, delete • Temporal: cease propagation if system clock is greater than date in configuration file (June 24, 2012) • Command and Control Server: upon activation on a new host, contact a command and control server (www.mypremierfutbol.com, www.todaysfutbol.com) via HTTP, [provides the opportunity track propagation and to modify or disable the software]	N/A

C. Analysis

After enumerating Stuxnet's effects and control features, we analyze these to determine if it complies with the principles of discrimination, distinction and proportionality.

1. Discrimination and Distinction

Although Stuxnet's propagation methods appear to be rather indiscriminant and lack distinction, its delivery of effects is neither indiscriminant nor lacking in distinction. Stuxnet sought to spread onto a wide range of Windows-based systems, presumably to increase the probability of reaching targets on closed networks. While the supervisory plane control features

provided some limits on the time frame and rate of propagation, Stuxnet was almost certain to propagate onto non-target systems, as was seen in its spread within Iran and beyond [15]. However, Stuxnet appeared to have only temporary, secondary effects on systems without the Siemens ICS software, taking no action beyond attempting to propagate. Conversely, Stuxnet's primary effects were applied with discrimination and distinction. The control features on the physical and logical planes limited delivery of primary effects to the specific combinations of ICS hardware and software suspected to be in use at the target facility and only these devices were functioning in a manner consistent with operating centrifuges for uranium enrichment. This combination of controls enabled Stuxnet to distinguish between targets and kept it from acting as an 'indiscriminant weapon.'

2. Proportionality

The possible collateral damage from Stuxnet's effects was in compliance with the principle of proportionality. Stuxnet was apparently designed to minimize collateral damage. Stuxnet affected only systems running ICS software with only those operating in very specific ways triggering the primary effects. Although there was a possibility of collateral damage to untargeted uranium enrichment facilities, the risk appears to be acceptable for the intended military objective.

As stated previously, we leave for others the question of the legitimacy of resorting to armed force to disrupt Iran's uranium enrichment operations. However, within the context of armed conflict, Stuxnet appears to have incorporated sufficient controls and targeting precision to represent a lawful application of force against this military objective.

5. CONCLUSION

It is apparent that operations in the cyber domain will grow in frequency and potential for collateral damage. Many questions remain regarding the legal issues of operations in the cyber domain and how to conduct these operations in a lawful manner. This paper has introduced a methodology for examining the targeting and control of cyber weapons and operations with respect to lawful armed conflict. This work is a step toward defining a common framework in which policy makers and personnel from the technical and legal disciplines examine these questions. Experience will no doubt enhance our understanding of this problem. It should also lead to better quantification of targets, effects and controls along with more formal processes for evaluation. Finally, the body of international law pertaining to armed conflict may expand to address questions of cyber weapons and operations.

REFERENCES:

[1] M. Schmitt, "Wired warfare: Computer network attack and jus in bello" International Review of the Red Cross, vol. 84, no. 846, June 2002.

[2] M. Sklerov, "Responding to International Cyber Attacks as Acts of War" in Inside Cyber Warfare, J. Carr. O'Reilly, 2009.

[3] T. Wingfield, "International Law and Information Operations" in Cyberpower and National Security, F. Kramer, S. Starr and L. Wentz eds. Potomac Press, 2009.

[4] M. Schmitt, “Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework” Columbia Journal of Transnational Law 37: 885, 913–15.
[5] N. Rowe, “The Ethics of Cyber War Attacks.” in Cyber War and Cyber Terrorism, A. Colarik and L. Janczewski, eds. The Idea Group, 2007.
[6] D. Kuehl, “From Cyberspace to Cyberpower” in Cyberpower and National Security, F. Kramer, S. Starr and L. Wentz eds. Potomac Press, 2009.
[7] U.S. Department of Defense, Department of Defense Strategy for Operating in Cyberspace. July, 2011.
[8] R. Clarke and R. Knake, Cyber War: The Next Threat to National Security and What to Do About It. Harper Collins, 2010.
[9] F. Kramer, “Cyberpower and National Security: Policy Recommendations for a Strategic Framework.” in Cyberpower and National Security, F. Kramer, S. Starr and L. Wentz eds. Potomac Press, 2009.
[10] G. Rattray, Strategic Warfare in Cyberspace. MIT Press, 2001.
[11] M. Swanson, et al., Contingency Planning Guide for Federal Information Systems, Special Publication 800-34 Rev. 1. National Institute of Standards and Technology: Gaithersburg, MD, 2010.
[12] U.S. Department of Defense, Joint Publication 3-13: Information Operations. February, 2006.
[13] J. Muir and P. van Oorschot. Internet Geolocation and Evasion, Technical Report TR-06-05, School of Computer Science, Carleton University, April, 2006.
[14] Q. Norton, “Anonymous 101.” Internet: http://www.wired.com/threatlevel/2011/11/anonymous-101. November 8, 2011 [February 1, 2012].
[15] N. Falliere, L. Murchu, and E. Chien, W32.Stuxnet Dossier. Symantec Corp., February, 2011.
[16] G. Brown, “Why Iran Didn’t Admit Stuxnet Was an Attack.” Joint Forces Quarterly, no. 63, October, 2011.
[17] L. Milevski, “Stuxnet and Strategy, A Special Operation in Cyberspace?” Joint Forces Quarterly, no. 63, October, 2011.
[18] S. Weinberger. “Is this the start of Cyberwarfare?” Nature, vol. 474, June, 2011.
[19] D. Albright, P. Brannan, and C. Walrond. Did Stuxnet Take Out 1,000 Centrifuges at the Natanz Enrichment Plant? Internet: http://isis-online.org/uploads/isis-reports/documents/stuxnet_FEP_22Dec2010.pdf, December, 2010 [February 1, 2012].
[20] P. Shakarian “Stuxnet: Cyberwar Revolution in Military Affairs.” Small Wars Journal, vol. 7, no. 4, April 2011.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Command and Control of Cyber Weapons

Enn Tyugu
Institute of Cybernetics
Tallinn University of Technology
Tallinn, Estonia
tyugu@ieee.org

Abstract: With the development of autonomous malware and autonomous anti-malware, command and control of cyber weapons is becoming an important part of cyber defence. In the present paper we discuss the dangers of deploying and controlling intelligent cyber weapons in a unified setting, considering these weapons as intelligent agents. Command and control of intelligent agents causes new threats that are difficult to avoid due to the complexity of behaviour of agents. Situation awareness of agents must be improved and verified, or at least carefully tested with respect to safety of their behaviour. Several possible dangerous behaviours of cyber weapons are discussed in the talk: misunderstanding of a situation, misinterpretation of commands, and loss of contact and formation of unwanted coalitions. A specific threat is the formation of unwanted coalitions by proactive weapons. This can happen if they get too much autonomy in decision making. A scenario of insubordination of agents is presented, considering a longer time perspective. General conclusions are the following: the more intelligent software becomes the more difficult it will be to control it; when designing and developing new cyber weapons, one has to guarantee the appropriate control over these weapons under any circumstances. It is practically impossible to use formal methods for verifying the safety of intelligent cyber weapons for their users. Setting strict constraints on the behaviour of cyber weapons and their careful testing are necessary.

Keywords: *command and control, intelligent cyber weapons, situation awareness, autonomous agents, proactiveness and adaptability in cyber defence*

1. INTRODUCTION

Command and control (C2) is a key aspect of any military activity, and according to a common understanding it concerns only human actors. With the development of autonomous malware and autonomous anti-malware, command and control of cyber weapons is becoming an important part of cyber defence. This is especially true for intelligent cyber weapons that can make decisions and autonomously plan actions. Hence, command and control must be extended to autonomous cyber weapons. An existing command and control application of this kind is known for botnets. However, it is still a simple case, because the botnets of today still have a

rather straightforward and simple way of operation. However, the situation changes when bots become more intelligent and get more freedom of action. Already in the foreseeable future we can expect much more proactive and intelligent cyber weapons both for offence and defence. One can classify them as intelligent agents and apply respective command and control. Special attention has to be paid to the cooperative behaviour of agents. In the long run, there exists the danger that intelligent agents may become too independent and they will perform unexpected and unwanted (harmful) actions. Avoiding this requires at least thorough verification of the possible behaviours of intelligent cyber weapons, and this is not a trivial task. For instance, on the phenomenological level one can easily postulate Asimov's laws of robotics, but to implement these laws requires more effort than one may expect.

A report from research firm Visiongain predicts that by the end of this year the cyber warfare market will be worth about sixteen billion dollars, as governments around the world invest further resources, creating new systems and protective measures to combat cyber criminals and hostile state hackers [1]. The Japanese newspaper Yomiuri Shimbun reported that the Defence Ministry's Technical Research and Development Institute began developing the anti-viral virus in 2008. Japan has reportedly requested for $2.3 million from Fujitsu to build a self-replicating assassin squad – a computer virus it can set loose in the network to track down and eliminate other viruses [2].

These are just examples, demonstrating that malicious software and cyber weapons are not only spreading, they are also becoming more sophisticated, independent and intelligent. In the present paper, we are analysing the possible consequences of deployment of powerful cyber weapons, in particular, the possibility of preserving control over these weapons. With the development of autonomous malware and autonomous anti-malware, command and control of cyber weapons is becoming an important part of cyber defence.

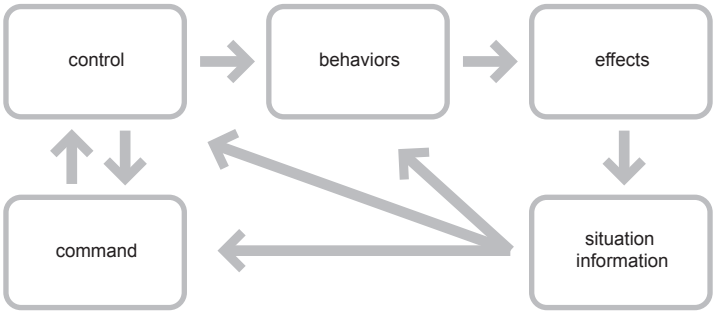
We compare, first, command and control as it has been understood in the context of military operations from one side, and in automatic cyber operations from the other side (Section 2). Then we introduce a generic concept of agent as an intelligent software component with proactivity (Section 3). We make some predictions about the further development of agents in cyber space, and describe their advanced features: beliefs-desires-intentions and reflection (Sections 4). We discuss threats of C2 of agents in Sections 5 and 6. Finally, we present a rather extreme scenario that may follow from the development of intelligence in agents. These scenarios may not become true, but are still possible in principle (Sections 7).

2. TWO FACES OF COMMAND AND COTROL

The magic words 'Network Centric Operations and Network Centric Warfare' point to the changed role of command and control in military operations. The changes concern, first of all, the speed of decision making and communication, but also the increased amount of information available for C2. A conceptual model of C2 is shown in Figure 1. The main control loop is as in control theory: control → behaviours → effects → situation information and back to control. The situation information is collected in classical control by sensors and is usually just a set of

values of variables. One cannot expect that this is the same in the present model. Understanding a situation in the context of military actions may be a very complex intellectual problem. The command and control actions are tightly bound by two-way communication links in this model.

FIGURE 1. CONCEPTUAL MODEL OF COMMAND AND CONTROL



An essential role in this model belongs to human factors as the list of C2 activities given in [3] shows:

- Establishing intent
- Determining roles, responsibilities, and relationships
- Establishing rules and constraints
- Monitoring and assessing the situation and progress
- Inspiring, motivating, and engendering trust
- Training and education
- Provisioning.

In cyber warfare, some command and control has been passed over to automatically operating entities – agents, command and control servers of botnets etc. – and this tendency is increasing. A good analogy can be found in air combat, where most actions are already performed completely automatically, and predictions for the next decade promise wide usage of artificial intelligence in the command and control loop. This increases the role of cognitive methods in situation awareness and situation management [4]. Changes are well visible in cooperative situation awareness of agents and humans. A crucial property is the speed of automatic C2 decision making. We will discuss these aspects in Section 5, dedicated to the threats of command and control of agents.

Let us look at the command and control in a case report of the Golden Cash botnet developed in 2008 and uncovered in 2009 [5]: "A user visits a legitimate, but compromised website which contains malicious Iframe. This Iframe causes the victim's browser to pull the exploit code from a server armed with the exploit toolkit. Upon successful exploitation, a special build of a Trojan, created for the attacker, is being pulled from Golden Cash server. Once installed, the Trojan reports back to the Golden Cash server and the attacker's account at Golden Cash is credited with currency. The first instruction sent by Golden Cash to the victim's machine, is to

install an FTP-grabber (to steal FTP-credentials) ... The victim's machine is now in a pool of infected machines controlled by Golden Cash and being auctioned to other criminals, using a different website for buyers ... The botnet's command and control server uses another website as a proxy that tunnels the bots communication to and from the C&C server. By applying this technique the C&C server remained 'protected' and undetected by security vendors for a longer time." Looking at the Golden Cash case, one can notice the following:

- automatic pay-per-install (including automatic pricing depending on the location of a buyer that varies from 5 USD to 100 USD per 1000 bots);
- automatic reuse of bots;
- information flow in two directions (from and to controller) to support the features above;
- usage of sophisticated malware products – Zeus and Zalupko Trojans;
- bots use FTP grabber to steal FTP credentials;
- using a proxy website by the C2 server.

The case of Golden Cash is over three years old. Considering threat predictions for 2012, we can see that the same botnet trade features are still dominating. Changes are in the architecture of botnets. Instead of a single centralised C2 server, peer-to-peer or hierarchical control is used. This requires more intelligent software and complex cooperation. Up-to-date information about botnet C2 servers can be found on the webpage of the Malware Threat Center of SRI International [6].

Botnets are used also on the defence side. A precedent has been created by the takedown of the Coreflood botnet in 2011 [7]. This takedown was authorised by the US DoJ and was performed by Internet Systems Consortium, Inc. (ISC) in cooperation with the FBI. It also demonstrates how simple it can be to change the side for C2 servers. The Coreflood servers were forced to talk to the FBI software, and shutdown commands were sent to infected computers. This required a Temporary Restraining Order (TRO) from a court.

3. AGENTS IN CYBER SPACE

The first ideas of organising software in the form of agents can be found in the actor model proposed by Carl Hewitt as a model of concurrent computations in the seventies [8]. This model has influenced even the development of object oriented languages. Today's agents can be considered, in essence, as well-developed objects that possess some features of intelligent behaviour.

Agents must have at least proactivity, the ability to communicate, and reactivity – the ability to make some decisions and to act. In software practice of today, agents are usually implemented on some special agent-based computing platform (cf. object-oriented software platforms). This simplifies the development and usage of agents, but it is not a necessary requirement. In the present paper we consider cyber space as an environment for agents, and we use a loose definition of agents as objects with the properties listed above. This is justified by the existing examples of malicious software that have agents' properties and move around in cyber space.

Cyber space requires some robustness and adaptability from agents, i.e. the ability to observe the environment and to use its features (protocols, operating system tables etc.). This is what characterises the advanced malware, and it is predictable that development will continue in this direction.

Probably the most sophisticated malware examples today that have agent properties are Stuxnet and Duqu [9]. They are very intelligent programmes (actually, a set of programmes) that analyse the environment in order to select a target, plan actions, are proactive and behave depending on time. Stuxnet consists of two parts: a delivery part that very selectively infects the control software, and a payload which is an intelligent and stealthy attacker of a special type of Siemens controllers. These parts can be considered as two autonomous agents.

On the other side, using intelligent agents in defence has been described in [10], where simulation shows that cooperating agents can effectively defend against DDoS attacks. After solving some legal [11] and also commercial problems, it should be possible, in principle, to develop a 'cyber police' consisting of mobile intelligent agents. This will require implementation of infrastructure for supporting the cyber agents' mobility and communication, but must be inaccessible for adversaries. This will require cooperation with ISPs. Multi-agent tools can provide a more complete operational picture of cyber space, for instance, a hybrid multi-agent and neural network-based intrusion detection method has been proposed [12]. Agent-based distributed intrusion detection is described in [13].

4. ADVANCED AGENT PROPERTIES

We have to look at some agents' features in order to be able to analyse the consequences of using agents as weapons (or as automatic warriors). These properties are reflection and beliefs-desires-intentions (BDI) – a combination of features that enable the agents to operate autonomously in a goal-oriented way. These are anthropomorphic features, and we must bear in mind that we should not apply any laws of human behaviour automatically to agents when considering these features.

A. Reflection

Reflection is the ability to perceive an agent's own state in the overall situation where an agent operates and to behave according to this perception, i.e. to use this for action planning. Reflection had already been introduced for objects in the eighties [14]. One can distinguish procedural reflection and declarative reflection. The first is implemented by programmes that have access to data describing the agent's/ object's state and, depending on the data, can change the functioning or even the programme of an agent or object (its behaviour in a more general setting).

Declarative reflection is the usage of models of environment and self for action planning [15]. Let us explain it in more detail. First, an agent must have a model that describes the current situation where the agent operates. It is important that this model includes as a part a model of the agent itself (this is the basis for a kind of consciousness that can appear in agents). Second, the agent must have a goal (or goals) presented by some data. Third, the agent must be able,

using these models, to plan its future actions for achieving the goals. In general, planning is a very challenging task. It can be simplified, when specific properties of the environment can be considered.

B. BDI and Emotions

The triplet of features belief-desire-intention got attention in psychology not too long ago, at the end of the last century, after M. Bratman presented his theory of human practical reasoning [16]. It also immediately got the attention of computer scientists for the programming of intelligent agents [17]. A project of the application of BDI in cyber defence has been described in [18].

The idea of BDI is to separate situation awareness from planning and execution of plans. The situation awareness is presented as beliefs – an agent 'believes' that the situation is as the agent 'sees' it. The desires represent a motivational state of an agent; they express the situations that an agent would like to achieve. Goals appear as a result of the analysis of the difference between the situation and the desired situation. Intentions appear as the goals that an agent decides to actively pursue. When a goal has been selected, a respective plan has to be obtained. This can be selected from a library of plans or it can be synthesised on the basis of existing information (beliefs). We present here an anthropomorphic explanation of BDI. Its software implementation is rather straightforward, using knowledge-based software technology. The most complicated part is planning. In the case of declarative reflection, plans are developed on the situation models. An example of planning for declarative reflection support is described in [15].

The steps from beliefs to desires and from desires to intentions depend on the emotional state of an individual or agent. We have not yet agreed on the presentation of emotions in agents. At present, we can speak about priorities instead of emotions. Handling priorities in computers is a common and well understood task.

One can expect that in the future a mechanism will be developed for controlling priorities that can be compared to emotions in human beings. The simplest model of emotions is as follows. Let us have a collection of priorities $p_1, p_2, ..., p_n$ that can control decision making in an agent: selection of goals, immediate reactions of an agent, interpretation of inputs etc. The number of priorities is large. Let us divide priorities into groups $e_1, e_2, ..., e_k$ in such a way that the priorities of one and the same group depend on a state s of the agent in a similar way. The number of groups is much less than the number of priorities: $k << n$. One can say that each group is controlled by an emotion. Thus we can define a small number of functions, $f_1(s), f_2(s), ..., f_k(s)$, for calculating a large number of priorities (a function f_i controls/ calculates priorities of the group e_i). This model can be extended by adding interactions between the groups.

5. THREATS OF AGENT COMMAND AND CONTROL

The agents have to be controlled by stating the most general goals and by giving some initial commands. Specific goals and a detailed action plan will be developed by agents themselves. However, the general command and control model shown in Figure 1 also applies to agents. It has links between its components responsible for command, control, behaviour and situation

awareness. In principle, any of these links can be attacked by an adversary. For instance, if it is true that the US RQ-170 Sentinel drone was captured by Iran [19], then this was obviously caused by an attack on the command and control system of the drone. It is argued that this was possibly done by the disturbance in the link between effects and situation information – wrong GPS data were passed to the control system of a drone.

The command and control of intelligent agents differs from C2 of botnets of today, because the agents have some independence. This makes their behaviour more difficult to predict, and this is a source of threats that can be:

- misinterpretation of commands;
- misunderstanding of a situation;
- unexpected emotions.

Misinterpretation of commands may be the most common threat, but it is also the easiest to avoid in principle. The threat appears if the language of C2, a communication protocol in the simplest case, is not sufficiently verified. Computer science supports verification of protocols, but it is still a complicated task. If an agent communication language is used which is more complicated than messages of a fixed format, then semantic problems of understanding appear. The language should be kept as simple as possible.

Misunderstanding of a situation can lead to wrong decisions at planning and execution stages. It is a threat that is difficult to avoid, because an agent operates in an environment that is complex or even unknown for the designers of the agent. The environment is cyber space, and it is complex with many different operating systems, software platforms, protocols etc. An obvious thing to do is to restrict the environment as much as possible by permitting the agent to operate only on known platforms. Situation awareness of agents must be improved and verified, or at least carefully tested with respect to the safety of their behaviour. A new trend is to apply artificial intelligence and cognitive methods in situation awareness [20]. This permits fusion of human and computer situation awareness and supports real time [21] and automatic [22] decision making.

The agents do not have emotions today, but they have to set priorities in order to be able to plan actions in a reasonable way – performing urgent and important actions first. A simple example of a mistake is setting a wrong priority on the basis of a false alarm. An analogy of a human activity is when someone fears that a threat exists and behaves in panic. It is a complex task to foresee all possible combinations that can appear in selecting priorities on the basis of the situation analysis.

6. MULTI-AGENT THREATS

Agents in cyber operations and cyber defence can be used most efficiently in multi-agent formations. Botnets could be an example, if bots are developed as agents. However, the control in botnets has still remained quite simple. Some cases of multi-agent defence are also available

from the literature [23,24]. One can expect that multi-agent systems will become the main form of agent application in cyber operations. In this case, agents will negotiate between themselves and will cooperatively create a complex behaviour for achieving the general goals stated by a commander. As a consequence, the strict control of behaviour of every single agent will be weaker. Also, it will be more difficult to foresee all possible cases for decision making. Practically, it will be impossible to verify the outcome of multi-agent behaviour for all situations. It is possible that backdoors and forced destruction will have to be built into agents. Multicast control messages will be needed for emergency cases of the agent control. Another option could be self-destruction of agents if loss of contact occurs, i.e. if for some time no command and control messages are received.

A specific threat of multi-agent systems is the formation of unwanted coalitions by agents. This can happen if agents get too much autonomy in decision making. Communication between the agents will be only partially observable to human controllers in this case. This will require very careful selection of constraints on the behaviour of agents. Here is the right place to remind of Asimov's laws for robots. This kind of law could improve the safety of multi-agent systems. However, there will never be an absolute guarantee of avoiding a misunderstanding of a situation by a team of agents. Also, a danger remains that a collection of agents may behave unintentionally in a harmful way. This is analysed in [25] and some possible, although not very probable, scenarios are described there. The next section presents one of these scenarios in a slightly modified way. This example should serve as a warning against neglecting the security of C2 of agents.

7. A SCARY SCENARIO

The year is 2030. Soon after the first attack, the Stuxnet malware was used in attacks on other systems developed by Siemens. It occurred to be a weapon applicable to various supervisory control and data acquisition (SCADA) systems, as soon as a system's design is known. Its intelligence was developed further with the aim of autonomously penetrating target systems. Its payload was adjusted to the target each time before launching.

Different cyber weapons were developed for performing different autonomous attacks. All these programmes can be called agents. They are quite autonomous, use BDI and declarative reflection, and can operate independently in an unfriendly environment.

As a consequence, botnets – the centrally controllable sets of passive programmes have evolved into armies of quite intelligent artificial fighters commanded in a net-centric way.

The intelligent malware caused much harm to the infrastructure of countries until multi-agent systems were also built for the defence. The defending agents were supported by advanced multi-agent platforms that gave them considerable advantage (in particular, good communication) compared with the attacking agents who had to operate in an unfriendly environment in a stealthy way. In order to further improve the capabilities of the defending agents, their autonomy was extended and their BDI system was developed more than ever before. This gave them an

excellent ability to plan their actions and even to set up new goals. This was very convenient for most of the users, and the general security awareness of people decreased to some extent.

The year is 2045. It was a bad idea to use too many agents with BDI. The danger was not so much in the intelligence of the weapons as in their willingness (and ability) to pursue their own goals. It became difficult to control very intelligent agents who had consciousness, priorities controlled by something similar to emotions, and who had their own desires.

A cyber conflict occurred between the agents that was initiated by the agents themselves. The country of the defending agents was immediately known, but the attacking agents seemed to belong to several different countries. It looked like there was a coalition of attacking agents from several countries. A lot of diplomacy was needed to clarify the case. A danger remains that agents may build hostile coalitions.

8. CONCLUDING REMARKS

The scenario presented above assumes the development of intelligent cyber weapons that are difficult to control. This is, in principle, a possible scenario. It is not based on any idea of artificial general intelligence (AGI) considered by the Singularity Institute for Artificial Intelligence in Palo Alto [26]. The AGI is based on an assumption that unsupervised learning capabilities of programmes will lead to an explosive growth in knowledge and intelligence of computers. Although possible in principle, and applied in data mining and parametric learning, the unsupervised learning has not developed to be applicable in learning on the conceptual level needed for understanding the world in general, and there are no signs of this possibility for the foreseeable future.

We have used the concept of agent for denoting a variety of cyber weapons of the future. This concept is used to denote just a set of features that provide autonomy, mobility and proactivity to the software under consideration. This has enabled us to analyse command and control of new cyber weapons in a unified setting, ignoring details of specific weapons. We have discussed the threats that are caused by agents, and we have made some unconventional predictions, assuming that the development of the cyber weapons will continue with acceleration. The future may not be as predicted here, but there is still good reason to be aware of the dangers described in the last sections of the paper.

We can point out some general conclusions. First of all, the more intelligent software becomes the more difficult it will be to control it. When designing and developing new cyber weapons, one has to be very cautious about guaranteeing the appropriate control over the weapons under any circumstances. It is practically impossible to use formal methods for verifying the safety of intelligent cyber weapons for their users. The global risks of wide implementation of artificial intelligence are analysed in [26].

One possible way to increase the safety seems to be imposing strict constraints on the behaviour of agents. This will be the analogy of the introduction of Asimov's laws on agents. However,

it will be still impossible to verify the correctness of behaviour of agents with respect to these constraints.

ACKNOWLEDGEMENTS

This research was supported by the Estonian Ministry of Education and Research target-financed research theme no. 0140007s12.

REFERENCES:

[1] *"The Cyberwarfare Market 2012-2022"*, [Online]. Available: http://www.visiongain.com/Report/732/The-Cyber-Security-Market-2012-2022, Dec. 05, 2011 [Feb. 7, 2012].

[2] J. A. Kaplan, *"Japan Reportedly Building Vigilante Virus Assassin Squad,"* Discovery News, [Online]. Available: http://news.discovery.com/tech/japan-vigilante-virus-120104.html, Jan. 4, 2012 [Feb. 11, 2012].

[3] David S. Alberts Richard E. Hayes, "Understanding Command And Control," *CCRP Publication Series*, DoD, [Online]. Available: www.dodccrp.org/files/Alberts_UC2.pdf [Feb. 7,2012].

[4] *Proc. IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2011, ISBN: 978-1-61284-784-9.

[5] E. Mills, *"Golden Cash botnet-leasing network uncovered,"* [Online]. Available: http://news.cnet.com/8301-1009_3-10266977-83.html, June 17, 2009 [Feb. 7, 2012].

[6] *"Most Prolific BotNet Command and Control Servers and Filters"* [Online]. Available: http://mtc.sri.com/live_data/cc_servers/, update daily [Feb. 07, 2012].

[7] S. Ragan, *"Coreflood: Botnet takedown introduces a potentially risky precedent"*, [Online]. Available: http://www.thetechherald.com/articles/Coreflood-Botnet-takedown-introduces-a-potentially-risky-precedent/13363/ Apr. 18, 2011 [Feb. 07, 2012].

[8] C. Hewitt, P. Bishop and R. Steiger, "A Universal Modular Actor Formalism for Artificial Intelligence," in *Proc. IJCAI*, 1973.

[9] R. Langer, "Stuxnet: Dissecting a Cyberwarfare Weapon", in *IEEE Security and Privacy, v.* 9, 2011, pp. 49 - 51.

[10] I. Kotenko, A. Konovalov, A. Shorov, "Agent-Based modelling and Simulation of Botnets and Botnet Defence," in *Proc. Conference on Cyber Conflict 2010*, C. Czosseck, K. Podins (Eds.), *CCD COE Publications*, Tallinn, Estonia, 2010.

[11] B. Stahl, D. Elizondo, M. Carroll-Mayer, Y. Zheng, K. Wakunuma, "Ethical and Legal Issues of the Use of Computational Intelligence Techniques in Computer Security and Computer Forensics," in *WCCI 2010 IEEE World Congress on Computational Intelligence, Barcelona, Spain*, 2010, pp. 1822 – 1829.

[12] E. Herrero, M. Corchado, A. Pellicer, A. Abraham,"Hybrid multiagent-neural network intrusion detection with mobile visualization," in *Innovations in Hybrid Intelligent Systems*, vol. 44, 2007, pp. 320–328.

[13] V. Chatzigiannakis, G. Androulidakis, B. Maglaris, *"A Distributed Intrusion Detection Prototype Using Security Agents,"* HP OpenView University Association, 2004.

[14] P. Maes, "Concepts and Experiments in Computational Reflection," in *Proc. OOPSLA*, 1987, pp. 147-155.

[15] M. Addibpur, E. Tyugu, "Declarative Reflection Tools For Agent Shells," in *Future Generation Computer Systems*, July 1996, pp. 1 - 12.

[16] M. Bratman. *Intention, Plans, and Practical Reason, CSLI Publications*. 1999.

[17] M. Georgeff, B. Pell, M. Pollack, M. Tambe, M. Wooldridge, "The Belief-Desire-Intention Model of Agency," in *Proceedings of the 5th International Workshop on Intelligent Agents V, Agent Theories, Architectures, and Languages*. Springer-Verlag, London, UK 1999.

[18] M. Shajari, A. Ghorbani, "Application of Belief-Desire-Intention Agents in Intrusion Detection & Response," in *Proc. Second Annual Conference on Privacy, Security and Trust*, 2004, Fredericton, NB E3B9W4, 2004, pp. 181 - 190.

[19] *"RQ-170 Sentinel Drone Downed In Iran Critical Updates,"* [Online]. Available: http://aviationintel.com/2011/12/08/downed-rq-170-sentinel-drone-critical-updates/, Dec. 8, 2011 [April 1, 2012].

[20] R. Jones, E. Connors, M. Endsley, "A Framework for Representing Agent and Human Situation Awareness," in *Proc. IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2011.

[21] T. Saarelainen, J. Timonen, "Tactical Management in Near Real-Time Systems," in *Proc. IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2011.
[22] Y. Fischer, A. Bauer, J. Beyerer, "A Conceptual Framework for Automatic Situation Assessment," in *Proc. IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2011.
[23] V. Gorodetski, I. Kotenko, "Multi-Agent Systems for Computer Network Security Assurance: Frameworks and Case Studies," in *Proceedings of the 2002 IEEE International Conference on Artificial Intelligence Systems (ICAIS'02)*, IEEE Computer Society, 2002.
[24] K. Boudaoud, Z. Guessoum, "A Multi-agents System for Network Security Management," in *SMARTNET '00 Proceedings of the IFIP TC6 WG6,7 Sixth International Conference on Intelligence in Networks: Telecommunication Network Intelligence*, Kluwer, The Netherlands, 2000.
[25] E. Tyugu, "When computers become too smart," in *Information Modelling and Knowledge Bases XXIII*, J. Henno, Jaak, Y. Kiyoki, T. Tokuda et al. (Eds.), IOS Press, Amsterdam, 2012, (Frontiers in Artificial Intelligence and Applications v. 237), 2012, pp. 412 - 418.
[26] E. Yudkowsky, "Artificial Intelligence as a Positive and Negative Factor in Global Risk," in *Global Catastrophic Risks*, N. Bostrom, M. Ćirković (Eds.), Oxford University Press, 2008, pp. 308–345.

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Case Study of the Miner Botnet

Daniel Plohmann
Cyber Defense Research Group
Fraunhofer FKIE
Wachtberg, Germany
daniel.plohmann@fkie.fraunhofer.de

Elmar Gerhards-Padilla
Cyber Defense Research Group
Fraunhofer FKIE
Wachtberg, Germany
elmar.gerhards-padilla@fkie.fraunhofer.de

Abstract: Malware and botnets are one of the most serious threats to today's Internet security. In this paper, we characterise the so-called "Miner Botnet". It received major media attention after massive distributed denial of service attacks against a wide range of German and Russian websites, mainly during August and September 2011. We use our insights on this botnet to outline current botnet-related money-making concepts and to show that multiple activities of this botnet are actually centred on the virtual anonymised currency Bitcoin, thus justifying the name.
Furthermore, we provide a binary-level analysis of the malware's design and components to illustrate the modularity of the previously mentioned concepts. We give an overview of the structure of the command-and-control protocol as well as of the botnet's architecture. Both centralised as well as distributed infrastructure aspects realised through peer-to-peer are present to run the botnet, the latter for increasing its resiliency. Finally, we provide the results of our ongoing tracking efforts that started in September 2011, focusing on the development of the botnet's size and geographic distribution. In addition we point out the challenge that is generally connected with size measurements of botnets due to the reachability of individual nodes and the persistence of IP addresses over time.

Keywords: *miner botnet, botnet analysis, cybercrime*

1. INTRODUCTION

Malicious software (short: malware) is the key enabler for digital crime and thus poses a serious threat to the modern society. One of its many uses is the creation of botnets. These networks of compromised computers (bots) are controlled by a third party (botmasters) and provide a flexible toolset for various illegal activities, promising remarkable financial gain with a low risk of being caught. Examples for activities are the massive sending of unsolicited messages (SPAM), distributed denial of service (DDoS) attacks, or the automated extraction of sensible credentials such as account login information or banking details.

One of the most recent botnet cases is the so-called "Miner botnet", named after its capabilities of mining Bitcoins. It received major media attention after carrying out massive DDoS attacks

against German websites (a detailed list is publicly available at [1]).
In this paper, we provide a comprehensive analysis of the "Miner botnet". Our contributions are the following:

- We analyse design and development aspects of a botnet on a technical level, covering individual binaries, the command & control (C&C) protocol, and its infrastructure.
- We present the results of our botnet tracking efforts since September 2011 and provide a statistical evaluation of the collected data set.
- We motivate current developments of botnet monetisation practices with one of the first specimens using the computational power of infected systems for direct profit generation.

The remainder of the paper is structured as follows. Section 2 covers background information and related work. Section 3 continues with insights on the botnet's infrastructure and outlines the characteristics of this malware specimen including monetisation of different functional aspects. Section 4 details the results of our botnet tracking efforts and Section 5 concludes this paper.

2. BACKGROUND

Centralised Botnets. The concept of botnets originates from the idea of enhancing malware with the ability to connect back to a server upon infection. First known cases of centralised botnets appeared in 1998/1999 and are tied to the so-called "Global Threat Bot" (GTBot), the remote access toolkit SubSeven and the email worm PrettyPark [2]. When infecting the target computer, these specimens joined a chat room on a specified Internet Relay Chat (IRC) server, notified the botmaster about their availability, and posted information gathered to enable further action. Obviously, the server's role in this centralised infrastructure is to provide C&C capabilities to the botmaster. The concept of using central servers evolved over the years, including masking of C&C servers through techniques like DNS Fast-Flux [3] and Domain Generation Algorithms (DGA) [4]. However, one flaw remains to this type of architecture from a botmaster's view: Shutting down all central C&C instances takes control away instantly and renders the botnet useless.

P2P Botnets. In order to overcome the drawback of depending on central components, experiments with peer-to-peer (P2P) mechanisms in malware date back as far as 2002 to the Slapper Worm [5]. The advantage of this technology is that the C&C channel is embedded into the botnet architecture, thus significantly contributing to resiliency against countermeasures when used correctly. A game-changing event was the appearance of the Nugache Worm, first detected in 2005 and considered to be responsible for the creation of one of the first botnets with a successfully distributed C&C infrastructure, based on a P2P protocol [6]. Since then, other P2P botnets have been observed and analysed. Detailed case studies have been performed e.g. for Storm [7], Waledac [8], and Conficker [9].

Bitcoin. Cybercriminals are constantly exploring new ways to generate profit from their botnets. Therefore, it was only a matter of time until bots were abused for generating Bitcoins

(BTC), an experimental digital currency scheme that was published in 2009 [10]. Bitcoins are calculated within a P2P network of competing nodes that iteratively perform SHA256 hashing operations towards certain target hashes. The first node to calculate an output hash based on certain input parameters that is below the target hash can claim a fixed amount of Bitcoins for its solution. The repeated hashing serves as a proof of work among competitors, who frequently join forces in so-called mining pools. Transactions of Bitcoins are cryptographically secured by a public-key infrastructure and the history of transactions is embedded into the calculations. While anonymity of transactions was not a design goal, techniques exist to aggravate tracing the flow of money. Bitcoins appeal to botmasters because they provide a way to immediately exploit the computational power of the compromised machines for financial gain. Bitcoins can be traded against hard currencies like USD or EUR on special trading platforms.

3. THE MINER BOTNET

In this section we present the characteristics of the Miner botnet. First, we provide chronological context of the operation of the Miner botnet. Next, we outline the development methodology used by the malware authors. We then focus our analysis on the set of executables specified by the botnet version number 1999. This version was the most recent on September 12, 2011 when we started our activities. The analysis is split by functionality aspects; for each we motivate the monetisation connected to it, namely:

- Pay-per-install (PPI) service for third parties
- Bitcoin mining
- Extortion via DDoS attacks
- Theft of social network identities

A. Timeline of Events

We were able to identify activities related to Miner back as far as December 20, 2010. On this day, a URL that can be linked to the botnet because of identical filenames was listed for the first time in the Abuse.ch Malware Database (AMaDa) [11]. Continuing our research, we concluded that at the beginning of this botnet, the malware was exclusively deployed and controlled via central servers using domain names of the following pattern: "<word>-<number>.ru", where <word> is a string e.g. "baza", "golos", "vn" and <number> an arbitrary number with two or three digits. Further related entries in AMaDa and investigation of binaries extracted from the botnet indicate it was mainly used for pay-per-install of adware and FakeAV in the first quarter of 2011. Beginning in March 2011, we found the distribution of a module for blocking access to the Russian social networks VKontakte.ru and Odnoklassniki.ru. We also identified the presence of an HTTP DDoS module since April 2011, but it is not known if the botnet was already used for attacks at this point. The first Bitcoin mining module appeared in late May/ early June 2011, at a time when mining became popular and Bitcoin calculation speed increased dramatically [12]. All of this information was gathered by comparing MD5 hashes of malware samples against their initial scan date on VirusTotal and other malware identification services available on the Internet. This type of botnet operation continued until July 2011, when the botnet infrastructure was migrated to a hybrid centralised/ P2P network as indicated in [13,14,15]. In August and September 2011, the Miner botnet carried out widespread DDoS

attacks against approximately 580 German websites. After September 17, 2011, only Russian websites have been targeted [1].

B. Botnet Topology and Command-and-Control Protocol

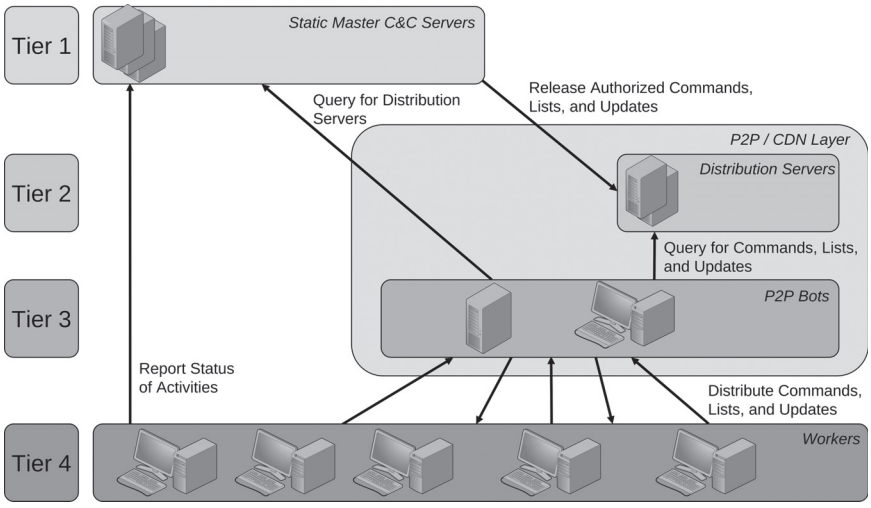
The topology of the Miner botnet can be divided into four tiers (Figure 1), sorted by descending relevance from the botmaster's perspective.

The top tier is represented by master C&C servers that are reachable through domain names and hard-coded IP addresses. These point to instances of nginx reverse proxy servers that are used to conceal the identity of the real C&C servers, which are operated by the botmasters and allow direct control and management of the botnet. Altogether, we identified less than 30 definitions of these fixed contact points. Most of them were not reachable during our investigations.

The second tier is defined through IP address lists published by the master C&C servers and consists of trusted peers that are internally called distribution servers. These servers are used to gather population information from the botnet, and to manage the connectivity as well as the flow of malware updates to the underlying tiers. They authorise and distribute lists for various purposes to third tier nodes. The third tier consists of all bots that are reachable from the public Internet and thus can be used as redistribution layer. In the following, we call nodes of this layer P2P bots.

The second and third tiers together form the P2P network of the Miner botnet. This network is primarily used as a malicious Content Delivery Network (CDN) and allows load balancing of binary transfers among its peers. It also serves as a backup layer for C&C in case the upper tiers are removed. The fourth layer consists of all remaining bots not reachable from the public Internet, e.g. because they reside in a private network. These bots serve as workers for operations like Bitcoin mining or DDoS attacks.

FIGURE 1. MINER BOTNET INFRASTRUCTURE.



The structure of the P2P communication protocol is shared by all tiers. The port used is fixed to 8080. In general, the protocol resembles HTTP GET-requests of the following structure "/search=<command{.txt}> HTTP/1.1". The URI path "/" and query variable "search" is static, while the actual command is appended as an argument. A query with the ".txt" extension serves as a status request and returns general information, e.g. the botnet version number or MD5 hash of contents to be transmitted by the actual command. For the full set of commands, see Table I. Answers to these requests have the structure of legitimate HTTP responses as generated by an nginx server, but are composed by the malware on the remote host.

TABLE I. P2P COMMAND&CONTROL PROTOCOL

Command	Answer (".txt")	Answer
error	-	returns an error code for the previously executed command
get_my_ip	0|<21 times "0">	returns the IP address as seen from the queried host
listen_test	0|<21 times "0">	requests the queried host to perform a connection check against local port 8081 in order to determine if the victim computer is reachable from the outside
test_server_r	0|<21 times "0">	this command is sent by a tier 3 node to a distribution server in order to validate if it is reachable from the public Internet
test_server	-	this command is induced by receiving the test_server_r command and performed by a distribution server
ip_list	0| <MD5 hash>	returns the most recent IP address list for subnet 1, the well-connected bots
Ip_list_2	0| <MD5 hash>	returns the most recent IP address list for subnet 2, the remaining bots
Ip_list_3	0| <MD5 hash>	returns an (empty) IP address list for the subnet 3
ddos_http_list	0| <MD5 hash>	returns a list of domain names to perform an HTTP-based DDoS attack against
ddos_udp_list	-	returns a list of domain names to perform a UDP-based DDoS attack against
btc_list	0 | <MD5 hash>	returns an IP address list for Bitcoin relay hosts
txt_server_list3	0 | <MD5 hash>	if the botnet is in fixed distribution mode, this command returns static download locations for the different modules
soft_list	<ver> | <MD5 hash>	returns a list of modules, each with its botnet version number, protected signature, filename and type ID
<filename>	<ver> | <signed MD5 hash>	returns the contents of the requested file

The communication protocol itself is not encrypted or obfuscated. For example, IP address lists are transmitted in plain text, with one quad-dotted IP address per line. The only mechanism of protection applied is a signature scheme for executable updates. Malware updates are delivered with an RSA-encrypted MD5 hash of the expected content. This is decrypted after downloading and checked against the actual MD5 hash calculated on the received data.

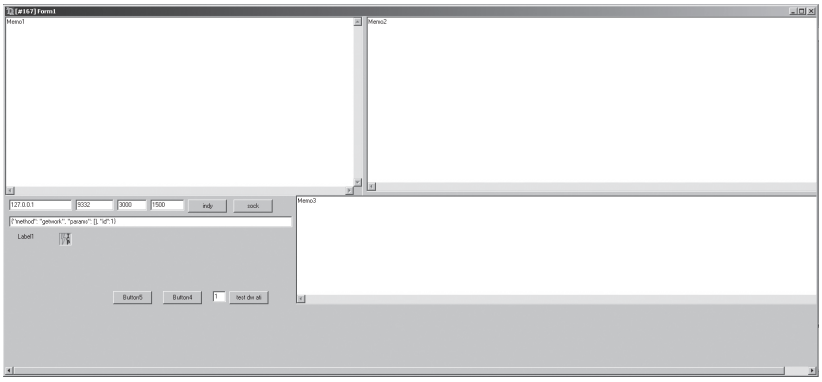
Besides the P2P communication, the static C&C servers and distribution servers are also contacted by bots on ports 80 and 62900 for submission of status reports and extracted data.

C. General Characteristics

Miner's code base can be characterised as a cluster of different malicious executables that are divided into modules according to their functionality. The functional groups identified can be categorised as infrastructure/ distribution, Bitcoin-related, DDoS-related, Social network-related, and additional utilities. The individual modules have no interdependencies. For example, the DDoS module can function on its own.

The dominant programming language used is Delphi version 2007. The majority of the malicious binaries have been developed by composing Delphi Graphical User Interfaces (GUI). Control elements such as memo and edit fields, labels, and buttons are placed on a form and assigned functionality, as shown in Figure 2. The overall behaviour is orchestrated by timers. While this approach may seem odd from a software development view, using a GUI allows easy live visual debugging of the modules by the malware authors. The visibility of the GUIs during runtime when deployed to victims is of course disabled so as not to raise immediate concern. The overall code base is structured into classes with different functional aspects, e.g. a class for nesting as a service, a class for refreshing IP address lists or a class for resolving its own IP address to geographical information. These classes are heavily reused among different modules.

FIGURE 2. A RUNTIME INVISIBLE GUI IS THE BASIS FOR FUNCTIONAL COMPOSITION OF SEVERAL MODULES, HERE FOR A BITCOIN MODULE (CLIENT_8.EXE). EXTRACTED WITH INTERACTIVE DELPHI RECONSTRUCTOR [16].



From 372 kilobytes (UDP DDoS module) to 1560 kilobytes (browser manipulation module), the footprint of the individual binaries is large compared with other malware. The reasons for the size are the mentioned graphical components and statically linked libraries. The authors make heavy use of third party open source products. The following libraries are present in all of the main modules: Internet Direct (Indy) [17] is used to implement communication interfaces and local web servers, Fast Giant Integers (FGInt) [18] supports implementation of a custom RSA signature scheme to protect malware updates, and RegExp Studio [19] allows matching of strings with regular expressions in various situations.

The binaries are not protected by any scheme that would harden them against analysis. Only

Ultimate Packer for eXecutables (UPX) is used in order to reduce the file size. The summarised file size of executables from botnet version 1999 is 17 MByte in decompressed state.

The only known spreading vector of Miner is social engineering of users through the social networks Facebook and VKontakte. As described in [14,15], the malware sends messages through stolen accounts to their friend lists and points users to fake YouTube videos that directly address the victim. In order to play the video, installation of a new Adobe Flash plugin is demanded, which is actually malware. The fake websites are directly hosted on P2P bots.

When a module is executed, it nests into the system in the following way. First, it copies itself to the Windows directory, either in the root directory or in a subfolder named "update.<number>" where <number> is a single digit number. The filename used for this purpose imitates typical Windows filenames (svchost.exe, svchostdriver.exe, sysdriver32.exe). It then restarts itself as a service and maintains presence on the system by enabling execution on system start-up. Malware configuration parameters are additionally stored in the Windows registry in module-specific subkeys that are also used for data sharing of timestamps or IP address lists. Most modules have a "close"-subkey that allows shutdown of the application through the registry by setting its value to "1".

D. Analysis of Botnet Version 1999

In the following, the different functionality aspects of Miner's malicious executable set of version 1999 are analysed. Each aspect is explained in context to the related binaries that represent the functionality. Furthermore the monetisation concept connected to the functionality is explained. We do not detail the changes between version 1999 and the current version 2103 because we did not identify major changes to the functionality.

1) Infrastructure/ Distribution

We first describe the core files responsible for integrating an infected machine into the botnet infrastructure. Next, we explain the different mechanisms of loading updates as well as SOCKS proxy and PPI as monetisation aspects.

loader2.exe – The first module to be executed on a freshly infected system is a loader that nests as a service called "srvsysdriver32" and then proceeds by performing an online connectivity test. If successful, it continues by contacting a random IP address from the embedded hard coded list of contact points, the so-called bootstrap list. As soon as a successful connection on port 62999 is established with a contact point, the loader continues by acquiring updated IP address lists of botnet peers with the commands "ip_list" and "ip_list_2". These first steps are similar for almost all modules. IP addresses from obtained lists are queried by the loader with the command "soft_list" in order to obtain the most recent list of modules. All files on this list are downloaded from different peers to establish full functionality on the infected system. Furthermore, a reachability test is performed with the command "listen_test" in order to determine whether the victim's computer can be accessed from the public Internet or not.

All of the downloaded files are first checked for a valid signature according to the implemented protection scheme. After successful signature validation, another check is performed against the modules type ID. If the type equals the ID of the distribution module and the reachability

test was positive, the node becomes a P2P bot, or else the module is not installed on the system, and the victim becomes a worker bot. All downloaded modules are registered to run on start-up and are executed to let them perform their initialisation.

The loader finally resolves its own IP to the corresponding country code and reports this information back on port 62900 to a list of master C&C servers, together with its module version number (1.66) and a unique identification number derived from the system's drive information and the computer name.

wdistrib.exe – The distribution module is the fundamental component of the flexible infrastructure of the Miner botnet. This module is only installed in case the machine is reachable over the public Internet. It does not install itself as a system service. When executed, hard-coded master C&C servers are contacted. Their authenticity is checked by comparing the content of a queried certificate against a fixed 13-digit number. After the authentication phase, a distribution level is queried from the server. This level decides whether a centralised or decentralised mechanism is used for distribution of malicious binaries. In either case, an IP address list of distribution servers is obtained. The entries of this list are then requested to ensure the own machine's reachability.

In the case of the centralised distribution level, a list of filenames is queried with the command "txt_server_list3" and the contents from the specified URLs are downloaded. The same is valid for current IP address lists and DDoS targets.

In the case of the two decentralised distribution levels, the value decides whether the victim's computer will be responsible for a network segment identified by an IP address list 1 or 2. From the list of distribution servers, a recent list of P2P bots from the chosen segment is requested. These bots have the same status as the victim's computer. The list is sequentially scanned for possible software updates, which allows injection of updates from any machine of the P2P layer. The refresh rate for IP address lists is set to 45 minutes. Independently from the distribution mode, downloaded files are offered to other infected machines that may contact the victim's computer.

Furthermore, a web server is opened with the purpose to serve a fake YouTube page that is used as the previously described spreading vector.

Lastly, a random port in the range of 10000 to 65000 is opened to serve as a SOCKS proxy service. This type of proxy is regularly used as an anonymisation mechanism and a well-known service in the cybercrime economy. The port number, IP address, country code, and result of a connection speed test performed against popular websites are reported back to the list of distribution servers. Depending on the speed test, nodes may be reassigned to subnet 1 or 2, the former containing the nodes that surpass a certain speed threshold.

This is also the first monetisation aspect of the Miner botnet, as the given information allows renting of compromised machines for the use as SOCKS proxy servers. We assume that the detailed information of country code and connection speed is used to justify individual pricing.

loader_rezerv.exe – This is a network-based downloader with the ability to install arbitrary executable files on a victim's computer. It contacts a range of hard-coded C&C URLs with the system identification number as well as its module version number (1.08). Upon connection, it can be commanded to download a file identified by a download ID from a given URL, together with the protection signature of the file. If download and validation are successful, a status message is reported back to the same C&C server.

gbot_loader.exe – The third loader has the malware to be spread directly embedded in its PE resource section. Upon execution, the system's geographical location is deferred and the payload is only installed when the IP address is associated with one of the following countries: USA, Canada, Australia, Great Britain, New Zealand, France, Germany, Sweden, The Netherlands, Italy, Belgium, Denmark, Swiss, Norway, Iraq, Israel, Qatar, Oman, Bahrain, or Japan. In the following, we refer to this module as the PPI module.

A detailed analysis of the observed payloads is out of the scope of this paper; therefore, we only give a short overview. In total, we extracted 11 different unique payloads from various gbot_loader.exe samples. All payloads differ massively from the binaries related to the Miner botnet. They are not written in Delphi and are protected against analysis. In 10 cases, a variant of the Max++/ZeroAccess rootkit was embedded. In one case we identified a variant of GBot/CycBot, a trojan downloader mainly connected to clickjacking.

The last two modules exemplify another monetisation aspect of the Miner botnet, pay-per-install. The presence of two separate mechanisms shows the importance of this feature to the botmasters. While loader_reserv.exe relies on the availability of the hard-coded domains and servers, gbot_loader.exe can and has been published through the Miner CDN to install third party malware.

2) Bitcoin-related Modules

The capability of Bitcoin mining is the most characteristic feature of the botnet and also responsible for its name.

btc_server.exe – This module is responsible for managing work distribution in the botnet and is only executed if the victim's computer fulfils the same properties as for the infrastructure distribution module. It serves as a proxy for the worker bots towards a selection of Bitcoin mining pools, clusters of miners that cooperate in order to increase their chance of gaining Bitcoins. It downloads one of the Bitcoin clients, namecoind or bitcoind, and joins a random mining pool chosen from a hard-coded list. These clients are used to backup the Bitcoin wallet containing earned Bitcoins. The wallet is posted every twenty minutes to a master C&C server. Furthermore, the module opens the ports 9442 and 9332 for Bitcoin communication with worker bots. Messages received by the workers are based on the Bitcoin JSON RPC protocol and delegated to the local Bitcoin client, which in turn forwards them to the chosen mining pool.

client_8.exe – This Bitcoin mining module is executed on bots of both tier 3 and tier 4. After nesting as service "srvbtcclient", a connection to the botnet is established and multiple operations are started in parallel.

Initially, an executable file named myunrar2.exe is downloaded from the Miner CDN. It works comparably to the utility geoip_unrar.exe and extracts the three embedded Bitcoin miners UFA Miner, RPCminer and Phoenix Miner. It then checks if the name of the video driver contains the string "radeon" and if so, checks for the driver revision installed. In case they are too old too perform Bitcoin mining on the Graphics Processing Unit (GPU) of the graphics card, the drivers are updated in the background through the vendor's website. After this, a speed test for the system is performed; the results including the system identifier, hardware information, the mining programme used and the hashing speed are submitted to a master C&C server. If an ATI graphics card is present, another test is executed on the GPU and the results augmented with detailed information about the graphics card are again posted to a master C&C server.

Further actions are the following. Every hour, a recent IP address list is obtained from tier 3 nodes. The IP addresses of this list are queried via a JSON RPC method for their current Bitcoin block count, which is returned if a btc_server.exe is running on queried node. In parallel, the successfully queried IP addresses are queried with another request for a portion of work. This allows the module to keep its own Bitcoin calculations at the current global network state. Finally, every five hours a status update about the mining operation is sent to a master C&C server.

The usage of Bitcoin mining on compromised machines is a remarkable development in botnets as it allows direct capitalisation through exploitation of computational power. While the Bitcoin currency has practically existed since early 2009, the first reports on malware used for stealing wallets of users were published in June 2011 [20]. After the Bitcoin exchange rate increased dramatically since late April 2011, with a peak of almost US$ 30 per Bitcoin in June 2011 [21], Bitcoin mining became economically justifiable for botmasters.

3) DDoS-related Modules

Next we explain the mechanisms of the DDoS modules and how the botmasters monitor attacked websites.

ddhttp.exe - The core module for DDoS attacks web servers via the HTTP protocol. It installs itself as a system service called "ddservice". After a connectivity check, it tries to download a list of DDoS targets. If the target list is acquired successfully, a status report with the unique system identifier and module version number (2.63) is sent to the contact point every 10 minutes. The following DDoS attack is performed with a randomly chosen User-Agent from a list of eight popular operating system and browser configurations. First, the IP addresses of the target host names are resolved and sanitised, i.e. the address 127.0.0.1 (localhost) is removed. Next, 10 concurrent threads are created to carry out the attack. In a first step, a connection to the target is tried in order to check if it is at all reachable. If the connection attempt is successful, the root path of the website is fetched. This page is then spidered for all links except RAR and ZIP archives, XLS, PDF, and DOC documents, executables, URLs containing "google.ru" or "cycounter" or email addresses. The attack then proceeds to request all the identified link targets to create even more load on the server.

udp.exe - The UDP DDoS component is a secondary attack module that was used during the massive attacks in August and September 2011. While the HTTP variant's goal is to exhaust the

web server application, the UDP module aims at saturating the network link of the target server. The module we analysed performs an attack against the hard-coded target "zenprotection.com", a DDoS protection service. It will send fragmented UDP packets with a payload size of 32001 identical, randomly chosen bytes to a random port in the range of 10 to 65000.

pele.exe - The last module related to DDoS allows the botmasters to evaluate the success of their attacks. Similarly to the HTTP module, it tries to obtain a list of currently attacked websites. It proceeds by requesting the root page of the attacked website and evaluates the HTTP status code. In cases where the status code indicates a redirection (3xx), the redirection is recursively resolved and queried.

Besides, a WHOIS lookup is performed to identify the "netname", i.e. the hosting service responsible for this domain. Next, the gathered data from these status checks is conducted into a report and submitted via a HTTP POST method to the master C&C servers. This procedure is repeated every two minutes. By inspecting the aggregated information from all reporting bots, the botmasters receive an almost real-time impression of how the attacked websites are reacting to the attack, and the detection of redirects or changes in reported features also allows them to adapt to countermeasures taken by their victims.

The monetisation aspect related to the DDoS functionality is extortion. While the first attacks were not connected to publicly known demands, following attacks were accompanied by emails from randomly generated yahoomail addresses requesting a payment of 100 Bitcoins to different account numbers for each target. We inspected eight Bitcoin account numbers identified through Google searches. According to the lookup service BlockExplorer [22], none of the accounts received incoming payment. The statement of Bitcoin as the desired payment method again underlines how firmly the botmasters pursue this virtual currency.

4) Social Network-related

The spreading success of the Miner Botnet in summer 2011 was heavily driven by the ability to interact with the social networks Facebook.com and VKontakte.ru.

iecheck12.exe – This module changes the Windows hosts file for static resolution of domain names with the intention to reroute the victim to a local proxy server when Facebook or VKontakte is accessed. This local proxy allows arbitrary interception and manipulation of website contents. On start-up, it queries current JavaScript files that are used to adjust the appearance of the social networks' websites to the malware's needs. Additionally, it downloads spam templates to be used for spreading.

Furthermore, the Geo IP database is used to determine the country the victim is situated in. Interestingly, the module also carries out the same functionality as the loader l_rezerv.exe by polling a list of hard-coded C&C servers for additional executables to download and install.

The core functionality of the module activates as soon as a victim logs into one of the mentioned social networks. The credentials are recorded and stored in the registry for multiple purposes. First, the credentials consisting of email address and passwords together with the system's unique identifier and geolocation are reported back to C&C servers. Next, the credentials are

abused in order to initiate communications based on the downloaded spam templates with individuals from the victim's friend list. These communications are not visible to the victim. The goal of the communications is tricking the contacted person via social engineering to download and install malware. In terms of monetisation, this provides the botmasters with stolen identities that serve as a tradable good.

Further investigation of the binary revealed the presence of a telephone number that appears to be connected to another fraud scheme. The telephone number appeared in multiple forums of Russian language where users reported that a popup blocked their access to VKontakte, demanding for a payment to obtain an unlock code via the Russian mobile service MTS. The forum entries date to January and February 2011 and might give a hint of the early uses of the botnet. However, we were not able to reproduce the mentioned functionality.

Analysis of further files received by the module revealed a scamming scheme based on the injection of an advertisement for a fake Groupon offer. In the advertisement, a payment of US$ 200 via PayPal is offered in exchange for a transfer of 25 Bitcoins to a given account number. We checked the addresses on BlockExplorer and did not notice any payments to the Bitcoin account.

5) Utilities
Miner makes use of two additional executables that support functionality to prolong its presence on the infected system. Both are only used once when initially infecting the system.

geoip_unrar.exe – This module uses the RAR archive algorithm to decompress an embedded Geo-IP database that allows derivation of a geolocation from an IP address. The structure of the Geo-IP database used by the Miner Botnet is similar to a reduced version of the free IP address to country database available from MaxMind [23].

resetr.exe – In order to reduce the chance of being detected or removed from the system, this utility disables and deletes the services responsible for Windows Update functionality and removes the Microsoft Background Intelligent Transfer Service (BITS) that is used for the roll-out of said updates and is also used for signature loading functionality of Microsoft Security Essentials (MSE).

4. MONITORING THE MINER BOTNET

In this section we present the results of our monitoring operation for the Miner botnet. First, we explain the focus of our efforts and the methodology we used, followed by an analysis of the data gathered.

A. Focus and Methodology

The focus of our operation was to get insights into the population and activity of the Miner botnet. Based on our findings on the botnet infrastructure (cf. section 3.B), we concentrated our efforts on the P2P layer. Information about active peers, commands, and malware updates can all be observed on this layer.

Our approach is an adaption of the techniques that were used for the monitoring of other P2P botnets [7,8]. The general methodology applied is recursive enumeration, also known as crawling. Starting with a set of bootstrap nodes, each of the nodes is queried for IP addresses of its known peers. By collecting these IP addresses and repeating the procedure on the growing set, the network can be enumerated until no new IP addresses are observed.

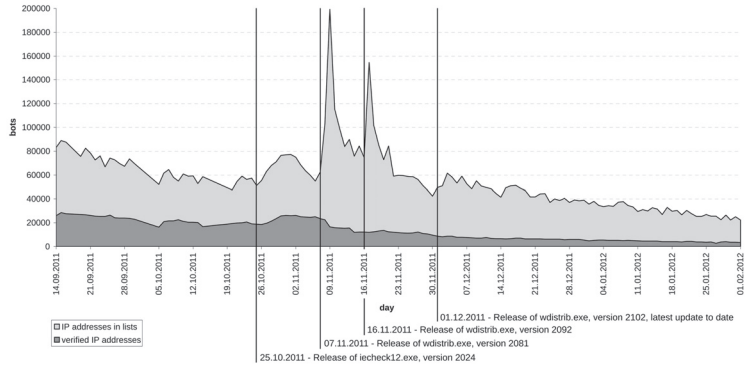
Applied to the Miner botnet, an initial bootstrap IP address list was extracted from the malware. The enumeration is done through the C&C protocol by using the command "ip_list" to query bots for their peer lists. The size of IP lists for segment 1 ranges between 80 and 250 and for segment 2 between 500 and 800 entries.

We created a tracking framework that implements this method and started crawling Miner's subnets 1 and 2 on an hourly basis on September 14, 2011. We chose an hourly interval according to the refresh rates set in the P2P bots (between 30 and 120 minutes) and in order to keep a low profile on the network. To gather additional information, we used the commands "soft_list" to identify files offered and "ddos_http_list" to obtain attack targets when a successful connection is established with a bot.

B. Results

For this analysis, we take the data gathered between September 14, 2011 and February 01, 2012 into concern. We present the data of subnets 1 and 2 combined, because we did not notice any discrepancy caused by the separation by network speed.

FIGURE 3. DAILY POPULATION OF THE MINER BOTNET.



We differentiate between the number of IP addresses observed through lists and the number of bots we were actually able to communicate with. On average, we could connect to 22.91% of the IP addresses listed, with a maximum of 46.26% on November 06, 2011. These percentages are mainly influenced by the embedded bootstrap lists and dynamics of bots joining or leaving the botnet, as well as the timeliness of IP address lists published by the distribution servers.

Figure 3 shows the development of the daily botnet population over time. We observed between 23,000 and nearly 200,000 peers in the IP lists and between 3,000 and 29,000 actually reachable

hosts. We have linked four remarkable events to activities in the botnet.

- On October 25, 2011, an update of the module for interfering with social networks was released. This caused a visible increase in infections.
- On November 07, 2011, updates to the PPI module and distribution module were published. The temporary spike in observed IP addresses in the lists can be explained with changes to the botnet backend. The actual decrease in population is probably caused by the updated PPI module. We assume that the botmasters of the Miner botnet sold a part of their population at this time.
- On November 16, 2011, the distribution module is updated again, causing another temporary spike.
- On December 01, 2011, an update to the distribution module was published, which was the last update to date. Since then, a constant decay is observable.

Table II shows that the geographical distribution of the botnet is centred on the countries Ukraine, Russia, Poland, Romania and Belarus. These countries made up about 70% of all infected hosts during our entire monitoring time. The only remarkable observation is that Poland and Romania change their position in November 2011. This is related to the above-mentioned event of an update to the PPI module.

TABLE II. GEOGRAPHICAL DISTRIBUTION OF REACHABLE BOTS

	14.09.11	**12.10.11**	**16.11.11**	**14.12.11**	**11.01.12**	**01.02.12**
Ukraine	28,24	29,49	22,27	25,88	25,29	27,04
Russia	18,56	20,76	17,11	19,35	21,74	17,83
Poland	10,40	9,15	10,55	7,25	7,24	8,06
Romania	8,97	7,77	12,07	11,44	11,71	9,98
Belarus	4,20	4,86	4,70	4,40	4,15	3,58
remaining	29,62	27,97	33,30	31,68	29,87	33,50

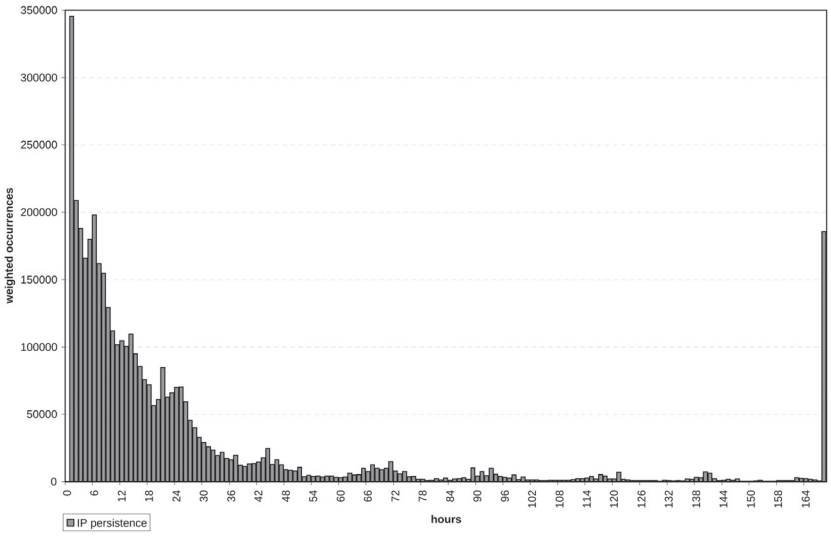
Furthermore, we analysed for how many hours single IP addresses were continuously present in the merged set of IP address lists. We took seven days from September 19, 2011 0:00 to September 25, 2011 23:59 as a sample, as shown in Figure 4. For possible intervals from 1 to 168 hours, the number of occurrences of an interval weighted with the number of hours it represents is shown.

The largest fraction is represented by short persistences with one hour occurrences being the most frequent. We assume this is at least partially influenced by the experience of a system infected with malware, especially in the case of Miner which increases system load through Bitcoin operations. The first significant drop-off is at about 6-8 hours which matches with the expected uptime of an office computer. The next decrease occurs after 24-25 hours and we conclude this is related to the enforced disconnect that many Internet Service Providers (ISPs) apply to their customers. The summarised weighted occurrences for 1 to 24 hours account for

72.43% of all occurrences, indicating that the majority of all observed persistences last for one day or shorter. This underlines that it is disputable to perform size measurements of botnets by counting observed IP addresses over longer time periods without taking the dynamics of the underlying systems and networks into concern, as has already been pointed out in [4].

The smaller peaks past the day mark are nearby multiples of 24 hours. We assume this to be caused by the way IP address lists are generated. The peak at 168 hours is caused by systems with a dedicated line and IP addresses that are constantly announced by the distribution servers. While having a strong impact in the representation chosen by us, these IP addresses account for less than 0.45% of all IP addresses observed in the given timeframe.

FIGURE 4. ANALYSIS OF IP PERSISTENCE IN THE WEEK FROM SEPTEMBER 19, 2011 TO SEPTEMBER 25, 2011.



5. CONCLUSION

In this paper, we have provided an overview of the Miner botnet. By taking this botnet as an example, we have motivated a selection of current techniques used by botmasters to extract money from their botnets. We outlined the chronological development of the botnet and its general characteristics. By this, it became obvious that the botnet owners have experimented with various methods for generating profits over time, adding and removing aspects, probably depending on how successful their activities were. We explained the layout of the hybrid infrastructure used in the botnet and detailed its capabilities and its C&C protocol. Furthermore, we presented our statistical data on its population and activities, gathered during four months of tracking efforts.

While the design and implementation used in this botnet are technically not on the same

level as of its more prominent competitors, the use of advanced concepts like a peer-to-peer infrastructure and RSA-signed updates indicate a trend that such features will become more and more common in all kinds of botnets in order to increase their resiliency against takedowns.

REFERENCES:

[1] F.Pfeiffer. *Minerbot Target List* [Online]. Available: http://www.ax10m.de/minerbot, Jan. 2, 2012 [Feb. 12, 2012].

[2] R. Ferguson. *The history of the botnet – Part 1* [Online]. Available: http://countermeasures.trendmicro.eu/the--history-of-the-botnet-part-i/, Sep. 24, 2010, [Feb. 12, 2012].

[3] The Honeynet Project & Research Alliance. "Know Your Enemy: Fast-Flux Service Networks," in *The Honeynet Project KYE Paper Series*, July 2007.

[4] B. S. Gross et al., "Your Botnet is My Botnet: Analysis of a Botnet Takeover" in *Proceedings of the 16th ACM conference on Computer and Communications Security*, 2009, pp. 635-647.

[5] D. Dittrich and S. Dietrich. "New Directions in Peer-to-Peer Malware," in *Sarnoff Symposium*, 2008, pp. 1-5.

[6] D. Dittrich and S. Dietrich. "P2P as botnet command and control: a deeper insight," in *Proceedings of the 3rd International Conference on Malicious and Unwanted Software*, 2008, pp.41-48.

[7] T. Holz et al., "Measurements and Mitigation of Peer-to-Peer-based Botnets: A Case Study on Storm Worm" in *Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats*, 2008.

[8] B. Stock et al., "Walowdac - Analysis of a Peer-to-Peer Botnet," in *Proceedings of the European Conference on Computer Network Defense*, 2009, pp. 13-20.

[9] F. Leder and T. Werner, "KYE: Containing Conficker", in *The Honeynet Project KYE Paper Series*, March 2009.

[10] S. Nakamoto. *Bitcoin: A Peer-to-Peer Electronic Cash System* [Online]. Available: http://www.bitcoin.org/bitcoin.pdf, May 24, 2009 [Feb. 2, 2012].

[11] abuse.ch. *abuse.ch Malware Database* [Online]. Available: http://amada.abuse.ch [Feb. 12, 2012].

[12] P. Wuille. *Bitcoin Charts* [Online]. Available: http://bitcoin.sipa.be/ [Feb. 12, 2012].

[13] T. Werner. *The Miner Botnet: Bitcoin Mining Goes Peer To Peer* [Online]. Available: http://www.securelist.com/en/blog/208193084/The_Miner_Botnet_Bitcoin_Mining_Goes_Peer_To_Peer, Aug. 19, 2011 [Feb. 12, 2012].

[14] R. Lipovsky. *Win32 DELF.QCZ: Trust Me, I'm Your Anti-Virus* [Online]. Available: http://blog.eset.com/2011/08/03/win32delf-qcztrust-me-i%E2%80%99m-your-anti-virus, Aug. 03, 2011 [Feb. 12, 2012].

[15] S. Duquette. *Win32 DELF.QCZ: Additional Details* [Online]. Available: http://blog.eset.com/2011/08/29/win32delf-qcz-additional-details, Aug. 29, 2011 [Feb. 12, 2012].

[16] *Interactive Delphi Reconstructor* [Online]. Available: http://kpnc.org/idr32/en [Feb. 12, 2012]

[17] *The Indy Project* [Online]. Available: http://www.indyproject.org [Feb. 12, 2012].

[18] Triade Systemm *Fast Gigantic Integers (FGInt)* [Online]. Available: http://www.submanifold.be/triade/GInt/gint.html [Feb. 12, 2012].

[19] A. Sorokin, *RegExp Studio* [Online]. Available: http://regexpstudio.com [Feb. 12, 2012].

[20] S. Doherty. *Infostealer.Coinbit* [Online]. Available: http://www.symantec.com/security_response/writeup.jsp?docid=2011-061615-3651-99 Jun. 16, 2011 [Feb. 12, 2012].

[21] *Bitcoin Charts* [Online]. Available: http://www.bitcoincharts.com [Feb. 12, 2012].

[22] *Bitcoin Block Explorer* [Online]. Available: http://www.blockexplorer.com [Feb. 12, 2012].

[23] MaxMind Inc. *GeoLite Country* [Online]. Available: http://geolite.maxmind.com/download/geoip/database/ [Feb. 12, 2012].

Chapter 6
Cyber Defence – Methods & Tools

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Natural Privacy Preservation Protocol for Electronic Mail

Kim Hartmann
Institute of Electronics, Signal Processing and Communication
Otto-von-Guericke University
Magdeburg, Germany
kim.hartmann@ovgu.de

Christoph Steup
Department of Distributed Systems
Otto-von-Guericke University
Magdeburg, Germany
steup@ivs.cs.ovgu.de

Abstract: Espionage plays a major role in military and paramilitary cyber warfare activities. While cyber espionage is mainly considered as the act of obtaining (confidential) information using illegal, technical methods, we have explored the possibilities of obtaining confidential material with technical methods using legal exploits. Due to routing conventions, messages containing confidential information may be sent through different states and herewith through conflicting communities. The servers that are used in this routing process are subject to the corresponding states legal system. In the case of electronic mail (e-mail), back-ups or copies being made are accessible to the corresponding authorities and/or private institutions. These copies of e-mails may be requested without knowledge of the sender or the sender's state and can be kept for an uncontrollable period of time. This may also heighten the risk of disclosure for encrypted messages.
We have developed a concept based on IPv6 to allow static and dynamic adjustment of the selected routes to maintain the specified or expected level of confidentiality. This concept may be developed further to be used as a privacy enhancing technology. The concept increases the level of control of transmitted data, technically enforces the expected or negotiated level of privacy and confidentiality, allows data tracking and heightens the user's awareness regarding the differences between postal and electronic mail.

Keywords: *privacy; confidentiality; IPv6; e-mail routing; IPv6 routing*

1. INTRODUCTION

Electronic mail (e-mail) is an important communication service based on TCP/IP and is sometimes said to be even more prominent than the World Wide Web. Especially in businesses, electronic mail is one of the major communication media used to transfer data and information [1].

As any communication medium that shall contact a preferably broad spectrum of individuals, e-mail needs to be easily accessible and hence needs to be able to transport data to almost any location. How this is done and the difficulties that arise due to this technique are described in the sections 2 and 3. At this point it is important to observe that the ubiquity accessibility of e-mail inherently implies a heightened exposure of the transferred data.

Since e-mail is mainly associated with a "point-to-point" communication between individuals or groups of individuals, users expect properties that are not given inherently. These properties are integrity, authenticity and a certain level of confidentiality. Different methods such as encryption and (digital) signing of e-mails aim at establishing one or more of these properties. However, these methods are barely being adopted by the broad public due to multiple reasons [2]. Some of the reasons may be the faulty assumption of a "point-to-point" communication or the association of electronic mail with postal mail and the expected legal implications, see section 2. Regardless what the reasons are, the result is that the majority of e-mails are transmitted without any guarantee of integrity, authenticity or confidentiality.

The small minority of e-mails that are protected by encryption and/or digital signatures are still at risk for manipulation and/or disclosure. Many encryption techniques rely on assumptions regarding the amount of time the protected data is exposed to attacks. It is commonly accepted, that it is rather difficult to hijack one specific e-mail. Expecting the common "man-in-the-middle"-scenario, where the attacker is a single individual without notable political or military power, this assumption may hold true.

However, if the standard "man-in-the-middle"-scenario may not be expected, as in the case of military or paramilitary cyber warfare activities, legal exploits may corrupt the idea of e-mails being hard to obtain. In fact, apart from espionage implications, e-mails may even be copied and preserved legally.

Current law, in most states of the European Union (EU), already demands the preservation of communication details for different periods of time, to allow the control of digital rights violations and for crime investigations. Obviously, this is not done for the purpose of espionage, but the technical practicability of the preservation of communication details and/or contents – without the knowledge of the involved communicating parties – must not be expected to be available to peaceful groups only.

We have developed a concept that provides the ability to influence the route selection and propose a model that additionally may take legal implications into consideration. This is either based on regional borders, "trusted parties" white-lists or information provided by the nodes involved in the e-mail transmission. Additionally to improving methods of control, our concept heightens the awareness regarding the differences between postal and electronic mail. The concept provides users with the ability to technically enforce a negotiated or expected level of confidentiality. We therefore believe that this concept may also be developed further to act as a new type of privacy enhancing technologies (PET).

The remaining part of this paper is structured as follows: The motivation for our work is given in the section 2 divided into legal and technical aspects of the problem described. Sections 3 and

4 give a short introduction to the prerequisites needed, while section 5 describes our concept. A concluding word and outlook is given in section 6.

2. A FAULTY ASSOCIATION

E-mail is often associated with postal mail and often explained as being the digital version of post. This association is faulty in both legal and technical terms and leads to the illusion of using a secured communication medium, see subsections 2.A and 2.B.

A. Technological Issues

Communicating through e-mail is commonly misinterpreted in two ways:

- Due to the association of e-mail with postal mail, the user expects an e-mail to be sent to the receiver directly. Since postal mail is commonly protected through national law and closed mail allows for a reasonable expectation of privacy, the user expects the company transporting and delivering the mail to be subject to national law. Hence, the delivery of e-mail is also expected to be done without the carrier opening, reading or copying and long-term storing the contents, the envelope or other information regarding the communication or the communicating parties.
- The user expects e-mail to be the digital version of a letter. Hence, the user visualizes the e-mail as an enveloped, closed, formally and directly addressed letter, transferred through national or international companies underlying strict laws, guaranteeing the privacy and/or secrecy of correspondence. A reasonable expectation of privacy is given.

1. Point-to-Point-Communication?

Users commonly believe that e-mail enables them to directly communicate with another individual or a selected group of individuals. Technically speaking, a "point-to-point" communication is expected, but not provided.

To allow e-mail to reach a user almost independently of the user's physical location and despite the offline/online-status a "live", point-to-point-communication may not be expected. Most individuals today understand e-mail as the digital version of postal mail and hence the e-mail folder is viewed as a digital "postal mailbox". This implies, that only the individual owning the folder and having access to it, may read the stored e-mail. Unfortunately, this assumption does not hold true.

While most e-mail providers will try to provide secure authentication methods to protect e-mail folders from being corrupted, legal implications may cause exceptions (cf. subsection 2.B).

The removal of an e-mail from the e-mail folder does not guarantee the removal of the e-mail from the server. Again, both technical issues (backup) as well as legal issues (data retention) may prohibit the deletion of contents for a certain amount of time.

Given that both the sender and receiver have e-mail folders and herewith e-mail servers that store the sent/received messages, there are already at least four parties involved in the communication.

However, the transfer of e-mail is mainly not done directly between the involved individuals mail-servers solely. This is mainly due to technical issues, as a) in most cases a direct route does not exist, b) routing conventions and c) traffic situations. The process of forwarding an e-mail through a network of nodes to one specific mail-server is called *e-mail routing* and will be explained further in section 3.

To guarantee the transmission of e-mail despite network difficulties precautions are made. Commonly, the nodes involved in the routing process will save a copy of the data prior to forwarding it. Normally, this data will only be stored for a short period of time. However, this "short period saving" is not guaranteed.

The selection of the route is done at the transport layer, i.e. based on IP routing conventions. These routing mechanisms take into account network and traffic parameters and chose the best (fastest and/or most reliable) route in technical terms. Whether state, company or legal borders are crossed is not part of the route selection process and mechanisms to provide this are currently not available. In fact, the route chosen is neither foreseeable nor evident to the users involved in the communication.

2. E-mail - An Enveloped Letter?

As e-mail is often associated with postal mail, the common visualization of an e-mail is an enveloped letter. This visualization has become so prominent, that the official symbol used by most mail-clients to display e-mails is an envelope.

However, sending a normal (unencrypted, unsigned) e-mail through the Internet should rather be associated with sending a postcard. An unencrypted, unsigned e-mail has neither protection nor guarantee of confidentiality. In this sense, e-mails are even less protected than postcards, as postcards commonly still remain protected by secrecy of correspondence laws.

As explained in subsection 2.A.1) an e-mail may not be expected to be transferred to the receiver directly, nor may it be expected that it is not copied or stored during its transmission. The regulations of the nodes depend on the node's location and the national law applying to them. This is not considered during route selection. No information about opening, copying or even routing of the e-mail is transferred to the user in a transparent or notable way. Moreover, users have currently no opportunity to influence the route selection.

B. Legal Issues

When using e-mail, the users involved assume the communication to be a point-to-point-communication between selected parties. This is neither the case for e-mail nor for postal mail. However, postal mail is commonly protected by secrecy of correspondence laws, guaranteeing that no-one except the sender, receiver and - in the certain, restricted cases – authorities may open or scan postal mail.

Due to the association between electronic and postal mail, the privacy regulations for postal mail are implicitly expected to naturally apply to e-mail as well. However, this assumption of the preservation of a "natural privacy" is currently not transferable to e-mail.

Current legal conventions within some states demand the preservation and surveillance of e-mail communication (with or without contents). In most cases, the communicating parties will not even be aware that their e-mail is being passed through another state due to routing conventions.

We claim, that a) there are differences in the understanding and the implications of secrecy of correspondence laws within different states and b) current laws already enforce the retention of communicated data (and communication related data) to different extents, providing different access options and storing the data for diverging periods of time.

Unfortunately, a complete analysis of the legal situation and understanding of the secrecy of correspondence can't be provided within this paper. To underline our claims, an impression of the understanding and implication of "secrecy of correspondence" is given for the United States and the European Union.

1. Secrecy of Correspondence in the U.S.

The Fourth Amendment to the United States Constitution is part of the Bill of Rights and protects citizens against "unreasonable searches and seizures". However, in the case of "probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons and things to be seized" exceptions are possible [3].

If and when the Fourth Amendment also protects digital data, has been discussed heavily. This ambiguousness is due to exceptions possible based on the "probable cause"-clause in the Fourth Amendment and the question if a reasonable expectation of privacy for electronic communication is feasible.

Two recent cases on the topic show the difficulties:

- In *Rehberg v. Paulk* (11.03.2010), the United States Court of Appeals for The Eleventh Circuit ruled that a person "does not have a reasonable expectation of privacy in an e-mail once any copy of the communication is delivered to a third party" [4].
- In *United States v. Warshak* (14.12.2010), the United States Court of Appeals for the Sixth Circuit ruled that a person ".. has a reasonable expectation of privacy in his emails.." and that the Fourth Amendment rights were violated by the government by compelling the Internet Service Provider (ISP) to provide access to e-mails "..without first obtaining a warrant based upon probable cause" [5].

Although the situation regarding the secrecy of correspondence is complicated in the U.S., there is no law enforcing ISPs to store and provide communication data as in the EU (cf. subsection 2.B.2). The Electronic Communications Privacy Act (ECPA) has been criticised by privacy advocates for not protecting all electronic communication and consumer's records. Moreover

it is claimed that the access to information stored at an ISP may be obtained too easily by governmental institutions.

In the United States Code, Title 18 – Crimes and Criminal Procedure, Part I – Crimes, Chapter 121 – Stored Wire and Electronic Communications and Transactional Records Access, § 2703, the requirements for the disclosure of costumer communications and records are given. Here it is stated, that the disclosure of data stored for less than 180 days is "only pursuant to a warrant issued using the procedures described in the Federal Rules of Criminal Procedure (or, in the case of a State court, issued using State warrant procedures) by a court of competent jurisdiction" [6]. Data stored for more than 180 days may be disclosed more easily. This also applies to (foreign) e-mails stored on servers in the U.S.

2. European Union and the Data Retention Directive

Most states of the EU have secrecy of correspondence laws, protected by the respective state's constitution. The secrecy of correspondence is explicitly protected by the European Convention on Human Rights, Article 8.

However, in March 2006 the Data Retention Directive was adopted by the EU. Members of the EU are required to store and provide the data specified within the directive for a period of at least 6 months, at most 24 months, for "the purpose of investigation, detection and prosecution of serious crime" [7,8].

The directive makes communication providers responsible for the gathering and storing of the required data. Affected by the directive are telephone, mobile telephone, internet access, e-mail and internet telephony communication data. The data stored must enable to identify the source and destination of the content transferred, the date, time, duration and type of communication as well as the device type used and - in the case of mobile communication - the location of the mobile equipment during the data transfer.

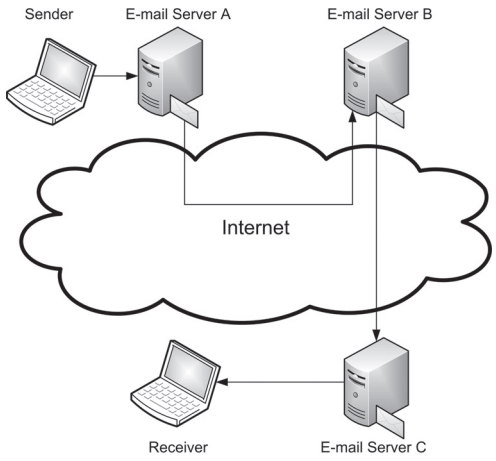
The member states were obligated to transfer the directive to national law until September 2007. Due to the existence of secrecy of correspondence laws within the member states, the implementation of the directive in national law was partially protested heavily. Currently 22 member states tried to transfer the directive to national law [9]. Sweden recently decided to postpone the decision. Romania, Germany and the Czech Republic had previously converted the directive to national law, but their respective courts ruled the directive to be unconstitutional. In 2010, the Irish High Court decided to challenge the Data Retention Law at the European Court of Justice. This was decided due to the previous juridical activities of the civil liberties campaign group "Digital Rights Ireland" (DRI) [10].

3. E-MAIL ROUTING

Sending an e-mail is done at the application layer of the OSI-Model. Common protocols involved in the process of sending and delivering e-mails are SMTP, POP3 and IMAP and their extensions. Relevant is, that all these protocols rely on the previous (i.e. lower) layers of the OSI-Stack to allow correct routing.

Fig. 1 shows how an e-mail is assumed to be forwarded. The sender of the e-mail composes the e-mail using a mail client or - more generally speaking - a Mail User Agent (MUA) and transfers it, by sending the e-mail to the senders mail-server. Here the local Mail Submission Agent (MSA) receives the message, looks-up the destination of the e-mail and forwards the message to the receiver.

FIGURE 1. BASIC VIEW ON E-MAIL COMMUNICATION FLOW



As mentioned previously, the routing through the Internet involves further parties and also the "look-up" of the receiver's MSA must be specified more correctly. Figure 2 shows a more detailed description of the routing process. As can be seen, the sender's MSA relies on the feedback given by a Domain Name System (DNS) server to resolve the receiver's domain name and determine the correct mail exchange (MX) server at the receiver's domain.

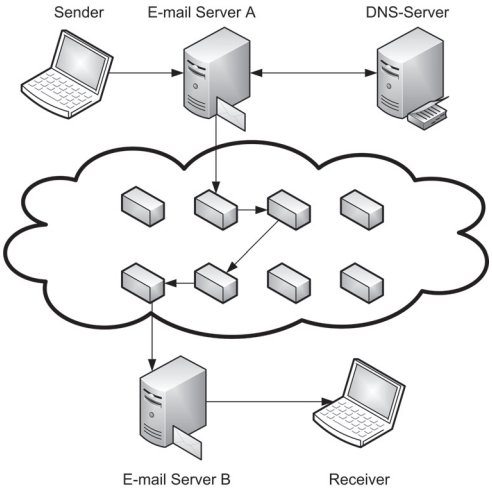
The DNS server responds with a list of MX records. The sender's MSA then transmits the message by passing it either directly to the correct Mail Transfer Agent (MTA), or by forwarding it through (several) MX servers closer to the destination.

It is clear that several servers are involved in the process of sending an e-mail, especially between different domains. Apart from the servers involved, one must also consider other network components (e.g. routers, gateways) needed to forward the message to the appropriate server.

The route selection is based on network properties, such as distance metrics. These metrics are internal properties and are acquired and used by routers without the user's knowledge.

To ensure e-mail delivery in case of network failures, nodes, such as routers, gateways and e-mail servers, store a copy of the e-mail locally to retransmit it if necessary. The lifetime of these copies is unspecified and depends on local factors such as memory usage or legal issues.

FIGURE 2. EXTENDED VIEW ON E-MAIL COMMUNICATION FLOW INCLUDING ROUTING AND DNS RESOLUTION



Given the numerous instances involved in the transmission of an e-mail, the fact that the nodes involved commonly store back-up copies of the IP-Packets and that the route selection is based on network properties, the vulnerability of e-mail communication becomes clear. If this vulnerability of e-mail transmission, the unguarded handling of e-mail by private and industrial users and the difficult and diverging legal situations are combined, the threats become apparent. Threats may be cyber espionage in general, industrial or military and paramilitary espionage. This may either be done by:

- Accessing the respective nodes involved in e-mail transmission. This may even be legal given some circumstances (cf. section 2.B),
- Attacking nodes known to be part of the routing of e-mails that are expected to carry useful information. Such nodes may be identified easily by finding the corresponding MTA(s) of a specific domain, e.g. a domain known to belong to military, governmental or similar institutions.

While MSA and MTA servers may only be guarded through intensive precautions done by the network administration, the route selection may be hardened through other concepts. The proposed method will allow controlling the transmission of data and preventing the transmission through untrusted nodes. This prevents both the usage of legal exploits as well as it may allow the selection of secure routes.

4. INTERNET PROTOCOL VERSION 6

IPv4 (Internet Protocol Version 4) [11] was standardized in 1981 when only 200 computers where interconnected. At this time, the defined address length of 32 bits was declared to be sufficient.

Due to the vast increase of computers connected to the Internet, the free IPv4-Addresses are about to be exhausted. This address exhaustion has been anticipated and threatened to have a limiting effect on the growth of the Internet. To prevent this limitation, a solution called IPv6 (Internet Protocol Version 6) was developed and published in 1998 [12]. This protocol extends the addresses to 128 bits, which is currently believed to be sufficient for some time.

The transition from IPv4 to IPv6 is ongoing, but the pace at which the transition is done is varying heavily, depending on the region [13]. However, it may be expected that eventually all nodes connected to the Internet will use IPv6.

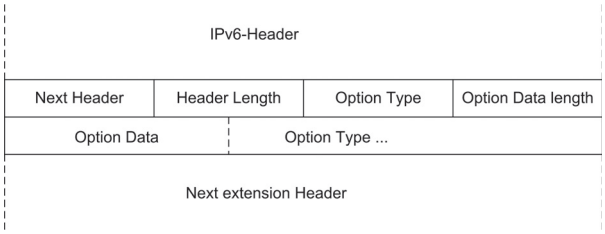
Besides providing a solution for the limited address space under IPv4, further improvements were made by IPv6 [14]. One of these improvements is the extensible header structure of IPv6. It consists of a base IPv6-Header and optional header extensions. This allows including additional transportation information.

A. Hop-by-Hop-Options

The information provided through the Hop-by-Hop-Options must be checked by each node along the route. Through these options, it is possible to pass further parameters and/or restrictions to the nodes processing the packet.

The Extension-Header contains a Next-Header-Field necessary for each IPv6-Extension-Header and a Header-Length-Field, to describe the basic layout. Hereafter an arbitrary number of options may be specified, by inputting a triplet of Options-Type, Options-Data-Length and Option-Data into the header.
The structure of an example Hop-By-Hop-Options-Header is depicted in Figure 3.

FIGURE 3. AN EXAMPLE HOP-BY-HOP-OPTIONS-HEADER



The uppermost three bits of the option type have a special meaning. The third bit defines the options data as unchangeable during transmission if set to 0. The other two bits describe what a node needs to do if the option type is unknown. There are four possible values:

00 – Skip the option
01 – Discard the whole packet
10 – Discard the packet and send ICMP error back
11 – Discard the packet and send ICMP error back if not multicast

Fortunately, the Hop-by-Hop-Options may be declared as mandatory. If a node cannot fulfill a specified option, the packed is dropped an error message is sent back to the previous node.

To ensure that the options are processed prior to the packet contents, the options should be placed directly behind the standard IPv6-Header-Fields. This is also recommended in the IPv6 RFC to restrict processing time in nodes.

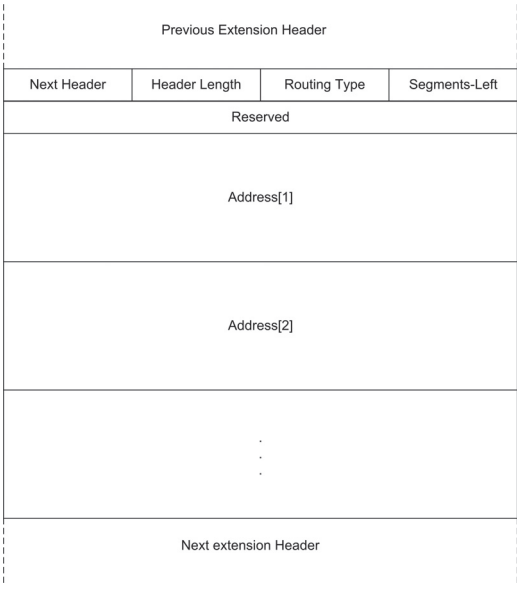
B. Routing-Options

The Routing-Options-Header enables users to specify the route a packet should take. The specified route may be “strict” or “loose”.

In the case of a “strict” route, the packet must be forwarded exactly along the nodes specified. If the route is declared “loose”, the given route may be considered as a recommendation.

The Routing-Options-Header keeps track of the next target in the route, by providing a counter, which is incremented by each node.An example of a Routing-Options-Header is depicted in Figure 4.

FIGURE 4. AN EXAMPLE OF A ROUTING-OPTIONS-HEADER



The described IPv6-Standard and the improvements made with the standard allowed us to develop a concept that may be integrated directly in the “backbone” of the Internet.

5. NATURAL PRIVACY PRESERVATION PROTOCOL (N3P)

The proposed Natural Privacy Preservation Protocol (N3P) is an extension of IPv6.

The IPv6 specification provides two approaches (cf. section 4) to influence the route selection at the network layer.

One method uses the Routing-Options-Header and is referred to as “Offline Route Selection” (cf. subsection 5.A), while the other method uses the Hop-by-Hop-Options and is called “Online Route Selection” (cf. subsection 5.B).

The offline route selection depends on a “white-list” of trustworthy nodes and needs to obtain this knowledge prior to the sending of data. The online method depends on the correct processing of the Hop-by-Hop-Options in each node, hence changes in the software of the nodes may be necessary. Hardware changes are only needed, if hardware modules to heighten the trustworthiness of a node are considered necessary. In both cases further work regarding the protection and trustworthiness of nodes should be done.

A. Offline Route Selection

The target of the offline route selection is to obtain a route prior to the message transmission, which is verified to satisfy the expected or needed level of privacy and to enforce exactly this route. The Hop-by-Hop-Options of the N3P-Packet are set, but may be omitted. This allows the offline route selection to use nodes that cannot process the Hop-by-Hop-Options, but that are considered as trustworthy by an authority.

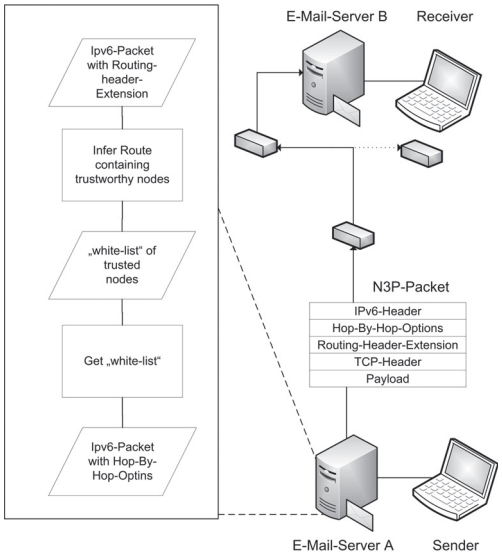
The mechanism can be divided into three major steps.

- First the “white-list” must be acquired.
- Secondly the route must be selected on which the message is to be transmitted.
- Finally the route must be attached to the packet in the IPv6-Routing-Header and the route must be marked as “strict”.

The basic steps can be seen in Figure 5.

Acquiring a trustworthy and correct “white-list” is quite complex but crucial. The “white-list” should be adjusted to the required level of confidentiality and is hence communication specific. Moreover, it must be protected against forgery and manipulation. This can be achieved through different solutions, for example digital signatures.

FIGURE 5. OVERVIEW OF THE OFFLINE-ROUTE-SELECTION



The route selection can be done in various ways. One simple solution is to "trace route" the target of the packet and check the resulting route against the white-list. However, this solution has the major drawback, that a failed check cannot be handled easily and needs additional mechanisms. Another approach is to obtain a topological view of the network through information provided by present routing protocols e.g. OSPF or EIGRP. This view may be utilised to run a routing algorithm locally and use the information as additional parameters. One example of a routing algorithm would be the Bellman-Ford-Algorithm with an additional confidentially-check for each node. If a node does not fulfil the confidentially requirements it is removed from the topology. After each node is checked, a route can be searched with the original algorithm. Other routing algorithms may also be modified to include the "white-list-checking".

Since the whole route selection and verification process is done at the sender, there is no need to modify intermediate nodes.

Two difficulties are:

- To obtain a trustworthy and up-to-date white-list
- The handling of node/network failures. As routing is done at the sender only, every node is a single point of failure on the static route.

An efficient solution is to partition the route into segments, where each segment must start and end at a trusted node. This makes the routing faster, as the amount of nodes that must be checked in each step is reduced. Additionally the effect of node failure on the overall routing is reduced.

B. Online Route Selection

If the route is not given prior to the transmission the routing process is called “Online Route Selection”. In this case, the selection of appropriate nodes is not decided prior to the sending of the data and “hard-coded” in the Routing-Header, but is instead done by the nodes along the route, chosen by the protocol specifications. This allows to abandon the usage of “white-lists” that may be difficult to obtain, synchronise and to keep updated.

To decide whether a node is sufficient or not in online route selection, two details must be known: The location of the node - to be able to derive legal implications - and the level of confidentiality a node can guarantee for a packet passed through that node.

While the first information (location) may be observed without the node cooperating, the latter is more complicated to obtain: The nodes involved must implement a method to decide and deliver their level of confidentiality. This implies that either the nodes must be trusted (without node adjustment) or further precautions to ensure the integrity of the method and its results must be taken (With node adjustment).

1. With Node Adjustment

To assess the “trustworthiness” of a node, reliable information about the nodes location must be obtained. Moreover, it must be guaranteed that a node is able to preserve the requested level of confidentiality (in terms of storage, processing, storage time, localisation etc.).
Since this information may not be obtained reliably without the node cooperating, we developed a 5-level confidentiality rating for nodes. Each node implementing the proposed method is assigned a confidentiality level, based on analysing the physical placement (legal issue) and the processing behaviour (information stored, period of time records are kept, data protection, etc.). The node’s level of confidentiality may be considered as a property of the node. A specific level may be demanded in the Hop-by-Hop-Options, declaring the minimum level needed to process the packet.

2. Without Node Adjustment

If the node adjustment is not possible, the option of gathering information about the location from neighboring nodes persists (for example DNS-Look-Up). The legal implications drawn from the physical placement of the node may then be used to derive an approximation of the level of confidentiality.

The DNS-Look-Up and the decision of forwarding the message to another node must be done by the current node. If this is applied recursively and consequently, a route of trusted nodes is selected. However, this implies that each node selected must be able to request the physical or network-based placement of the next node and evaluate its assumed confidentiality.

Independently of whether the confidentiality rating is done with or without node adjustment, the rating expected by the sender/receiver of a packet is placed in the Hop-by-Hop-Options-Field and marked with a type value of 0x84 to declare the options mandatory and unchangeable. The lower bits in the field represent an id of our specific option.

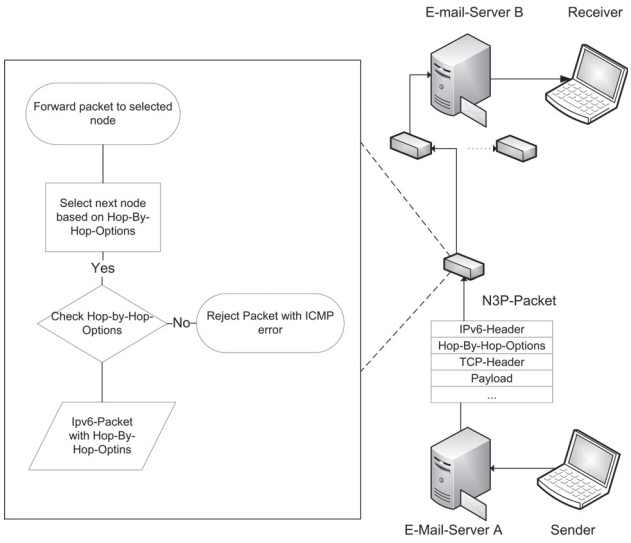
If a node does not fulfil the Hop-by-Hop-Options specified, either the node itself drops the packet (method "with node adjustment") or the current node may not forward the packet to the selected node (method "without node adjustment"). A basic view of the mechanism used by the online route selection can be seen in Figure 6.

The online route selection is more flexible and less restrictive than the offline route selection, however online selection implies larger implementation efforts.

In online selection, the decision of the confidentiality level of a node is crucial. As legal implications may influence confidentiality and depend on a node's location, some of our efforts are to automatically derive a value according to a node's placement. This value shall then be taken into consideration to determine the node's rating.

Other difficulties inherent to the online route selection are that concepts to ensure the trustworthiness of a node need to be implemented and an intelligent route selection procedure must be included to avoid looping.

FIGURE 6. OVERVIEW OF THE ONLINE-ROUTE-SELECTION



6. CONCLUSION AND OUTLOOK

Due to the faulty association of electronic and postal mail, governmental, entrepreneurial and private risk factors are induced by the misinterpretation of e-mail as a fast, reliable and secure communication medium.

It was shown that this is a misinterpretation in both legal and technical terms. The described exploit arose due to the combination of legal and technical issues. Interestingly, the complicated legal situation is both responsible for the unguarded use of e-mail and at the same time the reason for the implementation of techniques to silently copy and store transferred data. Although the intention is crime prevention and investigation, these techniques may also be adapted and used by other groups for lower purposes. This is especially the case if these groups do not consist of single criminal elements, but may rather be seen as military or paramilitary groups that have considerable influence on their local legislation.

Due to the described legal situation, especially within the EU, it is plausible to expect an increase of attacks on network structures known and demanded to save vulnerable data. Studies evaluating this assumption should be done. It should especially be evaluated if the amount of targeted attacks on infrastructures known to store vulnerable data increased since the integration of the EU Data Retention Directive.

The combination of technical issues and the legal situation in some countries leads to an uncontrollable amount of network nodes that silently may store transmitted data without restrictions. Access to the data is subject to the node's local government and may be legal, even if illegal in the sender's or the receiver's country. Both sender and receiver will not be aware of the fact that a copy of their communication may be processed outside their legal borders.

As copies of communication packets may be kept for an undefined and uncontrollable amount of time and the selected route is unknown, this may also comprise threats for encrypted messages. Encrypted messages may be exposed to surveillance and investigation without time limitations and may be compared with other encrypted/unencrypted messages from the sender.

Our method extends a solid, implemented, standardized and accepted protocol and herewith provides the ability to monitor, influence and control the IP routing. Since the new IPv6 standard is still being introduced, the proposed method may currently be adapted easily.

Two ways of extending/using the IPv6 standard were shown: The Routing-Header and the Hop-by-Hop-Options-Header. This provides the ability to either determine a static route prior to the message transmission or dynamically, e.g. "online" demand that only sufficient nodes may process a message and reporting must be done if the node is insufficient.

Difficulties in the offline/static route selection are to obtain a consistent and trustworthy "white-list" of nodes and to keep this up-to-date during transmission. However, since Internet connections have become more reliable and appropriate authorities exist, this problem may be expected to be solved rapidly. Further problems to investigate are the handling of manipulated nodes (list manipulation, route manipulation, spoofing, DDoS, etc.).

The difficulties arising in the online/dynamic route selection are a bit more complicated and wider:

- It must be ensured that the selected routes eventually allow the delivery of the transmitted messages.
- It must be guaranteed that the nodes either are trustworthy themselves or that trustworthy nodes possess the ability to gather information about other nodes before forwarding packets to them.
- An automated decision of a node's rating based on its location is under development.
- Implementation (hardware or software) of a node's confidentiality level rating method must be done and its reliability must be guaranteed.
- The proposed method should be tested with manipulated nodes to assess the potential effects.

Both the online and offline techniques have challenges, but also provide a unique gain of control to the e-mail routing process. In fact, the proposed concept may easily be transmitted to any type of communication based on IPv6.

Due to the control and feedback given to the user, our concept may be extended to a PET. Our concept entitles the user to influence the routing process and technically enforces the negotiated/expected level of privacy. The concept increases the user's awareness regarding the differences between electronic and postal mail and the resulting implications.

REFERENCES:

[1] Plantronics Inc. (2010). *How we work* [Online] Available: http://www.plantronics.com/us/howwework/

[2] S. Gaw et al.,"Secrecy, Flagging, and Paranoia: Adoption Criteria in Encrypted E-Mail," in Proc. of CHI 2006 Conference on Human Factors in Computing Systems, Montreal, Canada, 2006 © ACM Press, doi: 1-59593-372-7.

[3] T. Jefferson et al., *The Declaration of Independence and the Us Constitution with Bill of Rights & Amendments Plus the Articles of Confederation*. Washington, DC, Bottom of the Hill Publishing, 2010.

[4] J. L. Clerk, *"Charles A. Rehberg v. James V. Paulk," U.S. Court of Appeals for the eleventh circuit*, no. 09-11897, pp. *001-040* Mar, 2010.

[5] S. A. Spiegel, *"United States of America v. Warshak S.," U.S. Cout of Appeals for the sixth circuit*, no. 06-00111, pp. 001-098, Dec, 2010.

[6] Cornell University Law School. (2008, Jan, 8). *18 USC § 2703 – Required disclosure of customer communications or records*. [Online] Available: http://www.law.cornell.edu/uscode/text/18/2703.

[7] European Union, "DIRECTIVE 2006/24/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," *Official Journal of the European Union*, No. L 105, pp. 054-60, Mar, 2006

[8] The Register. (2005, Dec 14). *MEPs vote for mandatory data retention*. [Online] Available: http://www.theregister.co.uk/2005/12/14/eu_data_retention_vote/.

[9] European Commision. (2012, Jan 10). *European Commision Home Affairs*. [Online] Available: http://ec.europa.eu/home-affairs/policies/police/police_data_en.htm.

[10] Irish Times. (2010, May 5). *European cour to rule on storage law*. [Online] Available: http://www.irishtimes.com/newspaper/ireland/2010/0506/1224269793253.html.

[11] Information Sciences Institute University of Southern California. "INTERNET PROTOCOL DARPA INTERNET PROGRAM PROTOCOLSPECIFICATION," Defense Advanced Research Projects Agency, Arlington, RFC 791, 1981.

[12] S. Deering and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," The Internet Society, Reston, USA, RFC 2460, 1998.

[13] J. Prévost. (2003. Nov 10). *Deploying and using IPv6*. [Presentation]. Available: http://www.cu.ipv6tf.org/pdf/JP-DeployingIPv6.pdf.
[14] Dittler, H. P. *IPv6 das neue Internet-protokoll Technik – Anwendung –Migration*. 2. ed. Heidelberg, Germany: dpunkt.verlag, 2002

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Paradigm Change of Vehicle Cyber Security

Hiro Onishi
Alpine Electronics Research of America, Inc.
Strategic Research Group
Torrance, CA, USA
honishi@alpine-la.com

Abstract: Recently, cyber security for non-computers, such as transportation, utility, home appliance and others has become a serious social concern. Intelligent and electrificated modern vehicles have more MCU(micro controller unit)s, more software code than ever, which comes with huge cyber risks. Especially increased connectivity between vehicles and smart-phones / portable music-players changes the paradigm of vehicle cyber security, as virus and malware in smart-phones or music-players can invade automotive electronics. In this paper, first we introduce this new risk and assess the severity of this risk by a public risk assessment tool. Then we analyze the difficulties of cyber security in automotive electronics with limited network connectivity and low computational performance. Finally we conclude it with key findings and suggestions against this new risk.

Keywords: *cyber security, automotive electronics, vehicle connectivity, smart-phone, application download, DoS (Denial of Services)*

1. INTRODUCTION

Cyber security for computers has been discussed for a long time and many standards and guidelines have been published [1]. On the other hand, recently, cyber security for non-computers, such as transportation, utility, home appliances and others has become a serious social concern [2,3]. Even in automotive industry, from a long time ago, vehicles have large security risks, because they are expensive and frequently parked at unsecured locations. Besides illegally manipulated vehicles threaten drivers and passengers lives, and in the worst case, they damage communities in a large area [4,5]. Moreover current intelligent and electrificated modern vehicles have more MCU(micro controller unit)s, more software code than ever, which increases the risks to cyber attack [4,6,7]. Furthermore, increased standards or interoperability and common platforms or OS(operating system)s, such as, Windows™, LINUX™, AUTOSAR, GENIVI and others increase the cyber risks. Finally, "Road vehicle functional safety standard", ISO-26262 is raising the industrial concern about automotive electronics cyber risks [8].

2. EMERGING NEW VEHICLE CYBER RISK

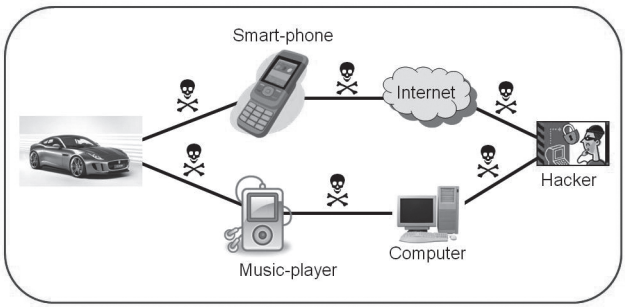
As modern vehicles have more convenient functions with wireless technologies, vehicle external connectivity increases cyber risks for automotive electronics [4,9,10]. At the initial phase of vehicle connectivity, GM OnStar, for example, mainly wireless communication modules installed within a vehicle, were used for emergency calls, concierge services, remote diagnosis and other automotive applications. However, recently, vehicle connectivity with carry-in devices, such as smart-phones, portable silicon music-players, portable GPS navigation systems, drive recorders and others is providing greater benefits to drivers. Table I shows recent factory-installed connectivity systems, which have been observed in "Los Angeles Autoshow - 2011". You can tell that under the red column systems, smart-phone has significant important roles, and under blue column systems, smart-phone that provides additional features or mobile phone connectivity is critical. As you can see in Table I, recently, mobile phones, especially smart-phones have more significant roles even in factory-installed connectivity systems. The growth of after-installation smart-phone connectivity system is obvious.

TABLE I. OEM(CAR MAKER)S' CONNECTIVITY SYSTEM (IN LOS ANGELE AUTOSHOW 2011)

OEM	System
Honda	(Telematics for Electric Vehicle) (USB smart-phone connection for CRV)
Toyota	entune™
Nissan	(Telematics for Electric Vehicle, to connect global data center)
Infinity	Infinity Connection® (for JX, by ATX)
Hyundai	blueLink®
Kia	UVO (Powered by Microsoft)
Ford	SYNC with MyFordTouch™
Lincoln	SYNC with MyLincolnTouch™
Cadillac	CUE (Cadillac User Experience)
Chevrolet	myLink
DCX	Uconnect
BMW	BMW ConnectedDrive

The growth of vehicle connectivity with carry-in devices is increasing vehicle cyber risk. FIGURE I shows the emerging vehicle cyber risks, caused by carry-in device connectivity. In this cyber risk, virus and malware attached with application software or music /video file, are first downloaded in carry-in devices. When carry-in devices are connected to the vehicles, virus and malware invade into the automotive electronics through vehicle entertainment systems or vehicle information terminals. In 2011 July, 82.2 million people in the US owned smart-phones [11]. Also, the number of application downloads on mobile phones is forecasted to reach 48 billion by 2015 [12]. Even now, many malware of Android™ OS smart-phone have been detected and they are increased by 472% from 2011 July to 2011 November [13]. Though this new type of cyber attack is not effective for the specified vehicles, this type of cyber attacks has become a critical threat for DoS (Denial of Service) for large number of unspecified vehicles, via anti-social activities.

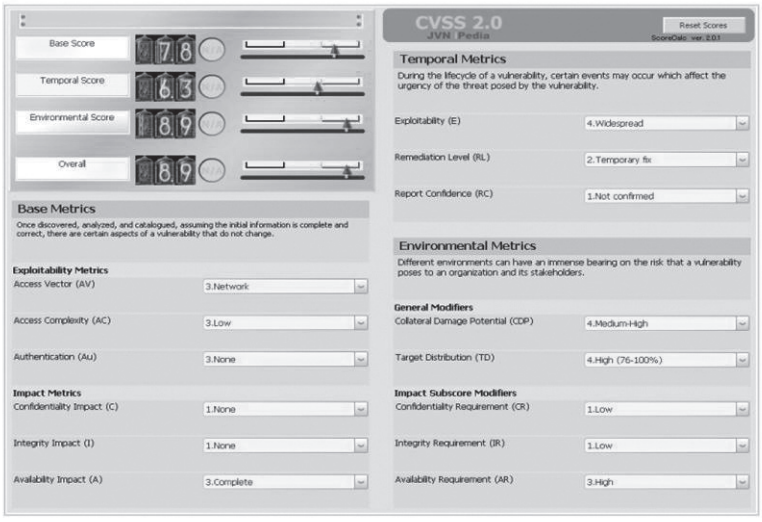
FIGURE 1. NEW VEHICLE CYBER RISK CAUSED BY CARRY-IN DEVICE CONNECTIVITY



3. ASSESSMENT OF NEW VEHICLE CYBER RISK

There are already many tools that can assess cyber security vulnerabilities. CVSS (Common Vulnerability Scoring System) calculator can score the cyber security vulnerability of systems or products with simple inputs and operations [14]. FIGURE II shows the vulnerability scores of the above mentioned new cyber risk by using this CVSS calculator. CVSS calculator assesses the highest severity-level of cyber risk as, "Level-3 (hazardous) - 8.9 of 10", due to its vulnerability against remote cyber attacks, lack of monitoring or protection mechanisms, wideness of damaged locations and the hazard of drivers, passengers or pedestrians lives.

FIGURE 2. OUTPUT OF CVSS (VERSION 2.0)
ABOUT CYBER RISKS CAUSED BY VEHICLE /CARRY-IN DEVICE CONNECTIVITY [14]



We have also estimated rough damages of this new emerging cyber risk with our assumption. First of all, 376,000 of one popular model vehicles were sold in the US and Canada, only

for year 2009. We assume that N [%] of these vehicles, i.e. (3,760 * N) vehicles are infected with virus. If we assume that 50% of these infected vehicles caused single-car-crashes and another 50% of these infected vehicles caused 2-cars-crashes, a total of (5,640 * N) vehicles are involved in crashes caused by this cyber risk. To more on, if we assume that average passenger number (including a driver) per vehicle is 1.5, a total of (8,460 * N) persons are involved in these crashes. If 50% of these (8,460 * N) persons would been killed or severely injured, the total number of fatalities or injuries would reach (4,230 * N). If we estimate an average of $10k financial damage per vehicle is involved in these crashes, including road facility damages (excluding fatality or injury damages), the total financial damage will reach ($56M * N). Table II shows these rough calculations based on our assumptions.

Table III shows the infection rates, N [%] vs. fatalities /injuries and financial damage estimations. Under the condition that N [%] is 1 [%], total number of fatalities and injuries becomes 4,230. This number is similarly equal to the total pedestrian traffic fatalities in the US per year (2008) and roughly 10% of all traffic fatalities in the US nationwide per year (2008) [15]. Besides, the total financial damage estimation reaches $56M, under the same condition.

TABLE II. ROUGH DAMAGE CALCULATION
(CAUSED BY VEHICLE /CARRY-IN DEVICE CONNECTIVITY)

Item	Assumption	Number	Notes / Reference
Target cars		376,000	One popular model sold per year (2009) in the US & Canada
Infection rate	$N_{[\%]}$		
Cyber-attacked vehicles	$N_{[\%]}$ are hacked	3,760 * N	
Cars involved crashes	50%: Single crash 50%: 2 cars crash	5,640 * N	
Persons involved crashes	Avg. 1.5 persons per car	8,460 * N	
Persons severely injured or killed	50%: Killed or Severely inured	4,230 * N	
Total damage cost	$10K per crashed car	$ 56million * N	Includes road facilities Excludes facilities & injuries

TABLE III. ROUGH DAMAGE ESTIMATION
(CAUSED BY VEHICLE /CARRY-IN DEVICE CONNECTIVITY)

Infection rate $N_{[\%]}$	Fatalities / Injuries*	Cost** [$ million]	Notes / Reference
0.01	42.3	.56	
0.1	423	5.6	
1	4,230	56	Pedestrian fatalities year ('08): 4,378 [15]
10	42,300	560	Traffic fatalities per year ('08): 37,261 [15]

*: Calculation is based on TABLE II.
Assumption: 50% of infected car have single car-crashes & the others have 2 cars-crashes
Average 1.5 passengers per vehicle, including a driver
50% passengers in crashes are killed or inured.
**: Calculation is based on TABLE II.
Assumption: Average $10 thousand damage per vehicle involved in crash,
including road facility damages, excluding fatality and medical cost

4. APPROACH FOR VEHICLE CYBER SECURITY

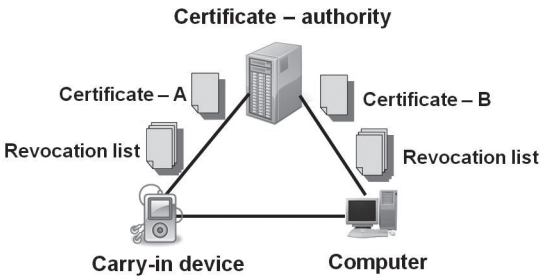
A. Reference: Approaches in Computer Cyber Security

In general, computer cyber security consists of encryption and certificate management. There are two types of encryptions, which are public key cryptosystem and common key cryptosystem [16]. Initially, public key cryptosystem, for example, RSA or DH (Diffie Hellman) is used to exchange small data, such as common keys in common key cryptosystem, for example, DES (Data Encryption Standards) or AES (Advanced Encryption Standards). Once a data sender and a data receiver can share with a common key, encrypted data with the common key can be exchanged between the sender and the receiver securely. Considering the balance between the required security level and durations of encryption and decryption (that depend on computational performance), a proper encryption algorithm is selected. Normally, current encryption algorithms cannot be broken within a reasonable time by existing ordinary computational performance [17].

A certificate-authority (also called as certification-anchor, certification-centre or trust-anchor) is monitoring whether a carry-in device is infected or under extraordinary conditions (FIGURE III). After a certificate-authority verifies a carry-in device condition, the certificate-authority can provide a certificate to the carry-in device without any issues. The carry-in devices with valid certificates can then connect to computers securely after a computer checks carry-in device certificates. In some security systems, a certification-authority distributes the revocation list, which includes names of carry-in devices with problems, so the revocation list can avoid the communication between carry-in devices with falsified certificates. This technique is called as "Remote (software) certification (=attestation) [18]. Secure boot is one different type of approach of certificate management system. It allows only signed software to run at the initial booting [19]. Though manufactures or system vendors cannot always track status of carry-in devices, because carry-in devices are connected at various locations, to various access points,

and to various usages, Thus, computer network can be protected against cyber attacks by using the certificate management system.

FIGURE 3. BASIC CONFIGURATION FOR COMPUTER CYBER SECURITY



B. US Government Initiatives

In the US federal government, mainly ICS-CERT(Industrial Control Systems Cyber Energy Response Team) in US DHS(Department of Homeland Security) is leading cyber security of industrial facilities, such as electric plants, electric-grids, water-lines and others, as well as all of the transportation systems, such as stations, trains, airport, airplanes, roads, bridges, vehicles, fleet and others. Recently US DOT(Department of Transportation) started cyber security activities in transportation areas. In August 2011, US DOT issued RFI(Request For Information) about vehicle cyber security, to collect information in this topic widely from automotive industry, IT industry, academia and others [20]. In December 2011, US DOT provided the first web seminar about cyber security, - "Introduction to Cyber Security Issues for Transportation"[3], and over 200 audiences joined it in real time. Besides, NHTSA(National Highway Transportation Safety Agency) of US DOT is strongly concerned about cyber security of automotive electronics [21]. Even TRB(Transportation Research Board) of NSF(National Science Foundation) established "Cyber security Subcommittee" under "Critical Transportation Infrastructure Protection Committee (committee number ABE40). This new subcommittee will cover cyber security for all transportation modes, such as aviation, airports, trains, rails, stations, transit, road infrastructure, vehicles, trucks, fleets and others, with communicating between other TRB committees or related US DOT organizations.

C. Key Players for Vehicle Cyber Security

Table IV shows government or public automotive research projects related with cyber security in the US and Europe. Right columns show security experts in each research project. As you can see, cyber security experts have already started research activities for the entire vehicle cyber security, such as vehicle-to-vehicle communication, MCU (Micro Controller Unit) protection and others [22-30].

TABLE IV. SECURITY PLAYERS FOR AUTOMOTIVE RESEARCH PROJECTS

Project	(major) region	Project leader & members	Security player(s)
Connected Vehicle Research [22] [23] [24] [25]		US Dept. of Transportation, 8OEMs(GM, HM, TYT, etc)	escrypt Embedded Security
CAR 2 CAR COMMUNICATION CONSORTIUM [26] [27]		(major OEMs, Tier-1 suppliers, etc)	
EVITA [28] evita & Preserve [29]		*EVITA:* BMW, Continental, Robert Bosch, etc	Architecture: escrypt Embedded Security Secure IC chip: Infineon, Fujitsu
Oversee [30]		VW, Fraunhofer, etc	escrypt Embedded Security

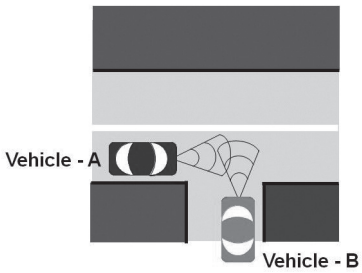
5. KEY FINDINGS AND SUGGESTIONS

A. Cyber Security Difficulties in Automotive Electronics

The certificate management system mentioned in the previous chapter can protect automotive electronics against ordinary cyber attacks, however new types or skilful virus or malware cannot be detected by a certificates-authority. In computer cyber security, virus or malware protection software is updated when a new virus or malware emerges. But, the first difficulty of automotive electronics is that online software updates have not prevailed yet, because of the limited vehicle external connectivity and risks caused by incomplete software updates [19].

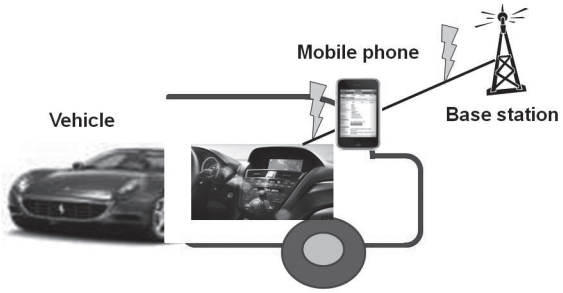
The second difficulty in vehicle cyber security is that automotive electronics have lower computational performance than ordinary computers, because of the high endurance (temperature, humidity, vibration and others) and longer vehicle life-cycle (over 10 years) compared to a computers' one (average 3 years). Then, in automotive electronics, old-generation MCU(Micro Controller Unit)s with low computational performance have to compete with hackers' latest-generation computers with high computational performance [4,31]. Therefore, cyber security, such as encryption or certificate management for automotive electronics has a higher risk to be broken in than ordinary computers' cyber security, because of the large computational performance difference between automotive electronics MCU(Micro Controller Unit)s and hackers' computers. Though secure encryption key storage is a very effective security method in ordinary computer cyber security, an encryption key has a higher risk to be stolen in automotive electronics, for the same reason. Once an encryption key is stolen, data inside or on the communication channels will be exposed. Furthermore, in the case that vehicles communicate with each other for crash avoidance (Figure 4), only limited encryption and certificate management are available, because of the time constrain (100 millisecond order). Due to the first and second difficulties, in general automotive electronics have higher risks to be infected than ordinary computers. Thus, counter measures for infected automotive electronics are more important than counter measures to avoid being infected, as compared to ordinary computer cyber security cases.

FIGURE 4. DIFFICULTY – (A) VEHICLE-TO-VEHICLE COMMUNICATION FOR CRASH AVOIDANCE



As for the third difficulty, the status of automotive electronics is more difficult to be monitored by a certificate-authority, as "Always-on connection" is not available yet. Especially, in the case if the vehicle can be connected externally only through a mobile phone (Figure 5). Once this mobile phone has been infected, this vehicle cannot receive diagnosis or treatment through the network. Though counter measures after infection are important in automotive electronics, a certificate authority cannot always monitor the status of automotive electronics, because of this difficulty. Therefore, in automotive electronics, the infection or extraordinary situation have to be detected within a vehicle. Another option is to trap virus or malware within a limited vehicle area, once a virus or malware enters in a vehicle to minimize the damages.

FIGURE 5. DIFFICULTY – (B) VEHICLE CONNECT THRU MOBILE-PHONE



In computer cyber security, DoS (Denial of Services) cyber risks can be reduced by treatment or isolates the infected computers, However as the last (forth) critical difficulty of automotive electronics, even if a small number of vehicles are infected, an infected vehicle can still threaten the drivers and passengers' lives. Because of this reason, even when automotive electronics are infected, vehicles safety should be maintained. Last but not the least, we should focus more on avoiding safety risks that threaten driver or passenger lives. In other words, we should analyse what happens when automotive electronics are infected and feedback these review results to vehicle designs.

B. Suggestions for Vehicle Cyber Security

FIGURE 6. SUGGESTED CONCEPT FOR VEHICLE CYBER SECURITY

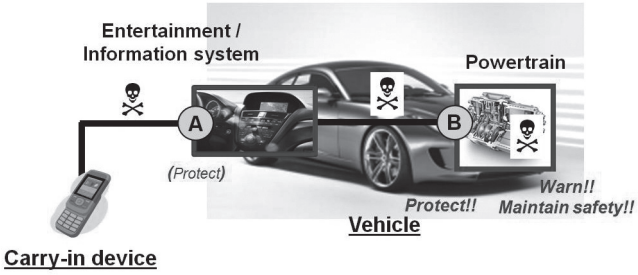


Figure 6 shows our basic suggestion with the consideration of the above mentioned difficulties. First, carry-in devices, such as smart-phones, portable music players or vulnerable mass-production computing devices, have high possibilities of being infected, because of various usages /applications, various places where they are been used and various access points. Therefore, virus or malware should be protected after point (A). However because of the limited computational performance and the limited connectivity of automotive electronics, it is difficult to protect against virus or malware at point (A), For these reasons, the first suggestion is to avoid virus or malware invasion into safety critical components or areas at point (B). One basic approach is to divide safety critical domains (areas, networks or components) from informative and convenient domains that have higher risks to be infected with more frequent external connectivity. Even if physical domain partitioning is difficult, logical partitioning, such as gateway insertion or virtual partitioning can be one of the approaches [4]. For the same purpose, hardware roles and software roles should be examined to avoid software manipulation caused by cyber attacks [19].

As for the second suggestion, even if virus or malware invade safety critical areas, it is very important to detect infection or abnormal condition quickly, and to inform them to the driver. This approach avoids critical accidents that threaten driver or passenger lives. So, the infection or extraordinary situation is supposed to be detected within a vehicle, because of the limited vehicle external connectivity. In other words, “Self-diagnosis”, “Self-detection” and “Self-warning” are more important. It is value that many automotive electronics devices or components are monitoring their individual status periodically and immediately warning drivers when something happen.

The last suggestion is to maintain safety even if safety critical components are infected. We should review what happens when automotive electronics are infected and feedback these review results to vehicle designs. As one example, when automotive electronics are infected, minimum fail-tolerance operations, such as, braking, stopping engines and opening the doors from inside, etc, are very effective to keep track. In this process, the concept of functional safety is very useful.

6. CONCLUSIONS AND NEXT STEPS

The growth of vehicle carry-in devices, such as smart-phones, portable silicon music-players and others are changing the paradigm of vehicle cyber risk. In the new emerging vehicle cyber risk, first, virus and malware are attached to applications or music /video file, and are downloaded to in carry-in devices, they then invade into automotive electronics (Figure 1). We assessed the vulnerability of this new emerging cyber risk by using a public cyber risk assessment tool (CVSS: Common Vulnerability Scoring System) [14] (FIGURE II), and also estimated the rough damages of this cyber risk based on our assumptions (Table II and Table III).

Comparing to ordinary computers, vehicle cyber security has many difficulties, such as "Limited connectivity", "Low computational performance" "Difficulty to monitor status of automotive electronics" and "Critical risk for drivers or passengers lives". As a consequence, counter measures after automotive electronics are infected, are more important than counter measures to avoid being infected. At the first plan, when virus or malware invade automotive electronics, safety critical components or areas, such as powertrain, braking and steering should be protected. Even if virus or malwares invade into safety critical areas, abnormal condition should be detected and be informed to a driver, quickly. Finally, when virus or malware invades into safety critical areas, at least, critical accidents that can threaten drivers or passengers' lives should be avoided.

In this paper, we have introduced risk analysis and problem findings. On a whole, as the next step, we are planning further the study on counter measures against this new cyber risk, and keep track with related governments initiatives, standards, researches and other activities worldwide.

REFERENCES:

[1] ISO "Information technology - Security techniques - Hash-functions" ISO/IEC 10118; ISO "Information technology - Security techniques - Key management" ISO/IEC 11770; ISO "Information technology - Security techniques - Trusted Platform Module" ISO/IEC 11889; ISO "Information technology - Security techniques - Evaluation criteria for IT security" ISO/IEC 15408 and others

[2] J. Cambridge *et al*., "Security and Critical Infrastructure Protection" TR NEWS, No. 275, Jul-August 2011

[3] M. Dinning *et al*., (2011, Dec, 7) "Introduction to Cyber Security Issues for Transportation" [Web seminar]. Available: www.pcb.its.dot.gov/t3/s111207/s111207_cybersecurity_intro.asp

[4] Information-Technolgy Promotion Agency (of Japanese government). (2011, Apr) "2010 report: Movements of Vehicle Cyber-security", (Japanese). Available: www.ipa.go.jp/security/fy22/reports/emb_car/documents/embsec_car2011.pdf

[5] K. Poulsen. (2010, Mar, 17). "Hacker Disables More Than 100 Cars Remotely" [Internet]. Available: www.wired.com/threatlevel/2010/03/hacker-bricks-cars/

[6] T. Kohno *et al*., "Experimental Security Analysis of a Modern Automobile" in IEEE Symposium on Security and Privacy 2010 [Internet]. Available: www.autosec.org/pubs/cars-oakland2010.pdf

[7] A. Weimerskirch, "Do Vehicles Need Data Security?" Society of Automotive Engineers World Congress, Detroit, MI, 2011

[8] ISO "Road vehicles - Functional safety" standard ISO 26262

[9] M. Raya, P. Papadimitratos and J.P. Hubaux, "SECURING VEHICULAR COMMUNICATIONS," Wireless Communications, IEEE , vol.13, no.5, pp.8-15, October 2006

[10] J.P. Hubaux, S. Capkun and J. Luo, "The security and privacy of smart vehicles," Security & Privacy, IEEE , vol.2, no.3, pp.49-55, May-June 2004

[11] (2011, Aug) “comScore Reports July 2011 U.S. Mobile Subscriber Market Share,” [Internet]. Available: http://www.comscore.com/Press_Events/Press_Releases/2011/8/comScore_Reports_July_2011_U.S._Mobile_Subscriber_Market_Share
[12] R. Vogelei, (2011, Jun) “Mobile Application Downloads to Approach 48 Billion in 2015,” [Internet]. Available: http://instat.com/press.asp?ID=3155&sku=IN1104930MCM
[13] E. Chickowski, (2011, Dec 1) “Android Mobile Security: A Growing Threat,” [Internet]. Available: http://mobile.channelinsider.com/c/a/Security/Android-Mobile-Security-A-Growing-Threat-548275/
[14] Information-Technology Promotion Agency (of Japanese government), “CVSS (Common Vulnerability Scoring System) Calculator”. Available: http://jvndb.jvn.jp/cvss/index.html
[15] U.S. Department of Transportation Research and Innovative Technology Administration Bureau of Transportation Statistics (2009) “Transportation Statistics Annual Report 2009”
[16] Information-Technology Promotion Agency (of Japanese government), (2008), “E-learning textbook about Cipher”, (Japanese). Available: www.ipa.go.jp/security/fy19/development/e_Learning_Cipher/index.html
[17] Encryption and Certificates, 1st ed., (Japanese), Nikkei BP publish, Tokyo, Japan
[18] Wikipedia, [Internet]. Available: http://en.wikipedia.org/wiki/Trusted_Computing
[19] A. Weimerskirch, “Security Considerations for Connected Vehicles,” in SAE Government/Industry Meeting, Washington DC, 2012 January. Available: http://www.sae.org/events/gim/presentations/2012/weimerskirch_escrypt.pdf
[20] Department of Transportation, “Cyber security and Safety of Motor Vehicles Equipped with Electronic Control Systems”, Solicitation Number: DTRT57-11-SS-00007, (2011, Aug, 2). Available: www.fbo.gov/index?s=opportunity&mode=form&id=40c0c2730b334df090dba322a61e956f&tab=core&_cview=0
[21] D. Smith, Opening Address of SAE Government/Industry Meeting, Washington DC, 2012 January.
[22] J. Sayer, ITS World Congress 2011, “Safety Pilot Model Deployment Test Conductor”, Orlando FL, 2011, Oct, 20
[23] Department of Transportation, “Safety Pilot Program Overview”, [Internet]. Available: www.its.dot.gov/safety_pilot/index.htm#6
[24] Department of Transportation, “Connected Vehicle Safety Pilot Program”, [Internet]. Available: www.its.dot.gov/factsheets/pdf/SafetyPilot_final.pdf
[25] V. Briggs, (2011, Aug, 3) “ITS Policy Program: Safety Policy Review and Discussion Introduction” [Web workshop]. Available: www.its.dot.gov/presentations/August_PolicyDay_v12_files/frame.htm
[26] A Weimerskirch, “V2X Security & Privacy: The Current State and Its Future” in ITS World Congress, Orlando, FL, 2011
[27] N. BiBmeyer, H. Stubing et al., “A Generic Public Key Infrastructure for Securing Car-to-X Communication” in ITS World Congress, Orlando, FL, 2011
[28] “EVITA(E-safety Vehicle Intrusion Protected Applications)”, [Internet]. Available: http://evita-project.org/
[29] “European R&D Project: PRESERVE(Preparing SecureVehicle-to-X communication systems)”, [Internet]. Available: http://cordis.europa.eu/fetch?CALLER=PROJ_ICT&ACTION=D&CAT=PROJ&RCN=97466
[30] “OVERSEE(Open Vehicular Secure Platform)”, [Internet]. Available: www.oversee-project.com/
[31] P. Kleberger, T. Olovsson and E. Jonsson, “Security aspects of the in-vehicle network in the connected car,” Intelligent Vehicles Symposium (IV), 2011 IEEE , vol., no., pp.528-533, 5-9 June 2011

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Sensing for Suspicion at Scale: A Bayesian Approach for Cyber Conflict Attribution and Reasoning

Harsha K. Kalutarage, Siraj A. Shaikh, Qin Zhou and Anne E. James
Digital Security and Forensics (SaFe) Research Group
Department of Computing
Faculty of Engineering and Computing
Coventry University
Coventry, CV1 5FB, UK
{kalutarh, aa8135, cex371, csx118}@coventry.ac.uk

Abstract: Cyber conflict monitoring remains one of the biggest challenges today, amidst increasing scaling up of cyberspace in terms of size, bandwidth and volume. Added to this, the increased determination of cyber actors to operate beneath the threshold makes it ever more difficult to identify unauthorised activities with desired levels of certainty and demonstrability. We acknowledge a case for persistent and pervasive monitoring; detection of serious sabotage and espionage activities, however, is dependent, in part, upon the ability to maintain traffic history over extended periods of time, somewhat beyond current computational and operational constraints. This makes it crucial for research in cyber monitoring infrastructures, which are configured to handle cyberspace at live and modern scale and sense suspicious activity for further investigation. This paper explores Bayesian methods together with statistical normality to judge for effective activity attribution, particularly in high-volume high-scale environments, by combining both prior and posterior knowledge in the scenario. The set of experiments presented in this paper provides tactical and operational principles for systematic and efficient profiling and attribution of activity. Such principles serve a useful purpose for technologists and policy-makers who want to monitor cyberspace for suspicious and malicious behaviour, and narrow down to likely sources. The proposed approach is domain agnostic and hence of interest to a cross-disciplinary audience interested in technology, policy and legal aspects of cyber defence.

Keywords: *anomaly detection, Bayesian approach, reasoning, attribution, cyber attacks*

1. INTRODUCTION

Cyber conflicts are increasingly a part of mainstream warfare. Attribution of cyber activity — "knowing who is attacking you" or "determining the identity or location of an attacker or an attacker's intermediary" [1,2,3] – is naturally a vital ingredient in any cyber security strategy. Parker claims that 'a common problem with many analysis tools and techniques today is that they are simply not designed for purposes of attribution' [4]. According to [5,6], although current approaches are capable of alerting to suspicious activities, they are failing in this information age because when computers are under attack, the 'who' and 'why' are frequently unknown. Many researchers claim that completely depending on information derived from network traces will do little for cyber conflict attribution and detection, mainly due to the nature of Internet infrastructure, and therefore there is a need for approaches that combine technical solutions data with the information gathered from contextual analysis and intelligent services (combining prior belief with posterior knowledge) [7]. This paper aims to address this challenge and presents work ultimately contributing towards this goal.

Reconnaissance, the first phase of the anatomy of a cyber attack, can be further sub-divided into three incremental stages: casing, scanning, and enumeration. It is difficult to tackle suspicious activities at the casing stage, as everything seems to be legitimate. However in the second stage, scanning, the attacker attempts to send packets to the target IP address (range of IP addresses) with the goal of determining what machines are presented and reachable (ports) on the target network. Two most common examples of scans, among many others, are 'pings-ICMP' and 'SYN-TCP'. This offers a starting-point for detection of potentially suspicious activity. For enumeration, the attacker may follow up with various kinds of attempts to identify services. The detection of scan and enumeration attempts is made more difficult as attackers increasingly use slow scan rates to stay beneath the threshold. If an attacker is methodical enough to make only the slightest of changes at any one time and each step is spaced far enough apart, it will be difficult to detect by traditional signature matching algorithms. Often, network-based intrusions signatures are state-full and require several pieces of data to match an attack signature. If the length of *event horizon* (time amount from the initial data piece to the final data piece needed to complete the attack signature) is longer, intrusion detection systems (IDSs) cannot maintain state information indefinitely without eventually running out of resources. This helps slow attackers to hide behind noise and other traffic. Most current approaches do not track activity over an extended period of time, due to computational constraints and disc storage requirements. This paper develops an approach to serve as an early warning system for slow suspicious activities that warrant further investigation.

This work is inspired by Chivers *et al.'s* work [8,9] to adopt a Bayesian approach to combine both prior and posterior knowledge in the scenario and detect (with attribution) slow and suspicious activities in a cyber conflict. The series of experiments examines the effectiveness of such an approach under different parameters: multiple attackers, traffic volume, cluster size and event sampling.

The rest of this paper is organised as follows: Section 2 presents a brief overview of related work; Section 3 presents the underlying methodology and the theoretical account of the process; Section 4 overviews the experimental set up and Section 5 follows up with results and analysis.

Section 6 presents some early results on possible use of sampling. Section 7 concludes the paper.

2. RELATED WORK

Although a considerable number of anomalies-based intrusion detection approaches have been proposed during the last two decades, many of them are general in nature and quite simple [10-12]. They fail in attributing, in accumulating evidence, and also in scaling up. Since our approach accumulates evidence (both contextual and technical traits) over an extended period of time and uses that information to identify aberrant behaviours (see Sections 3 and 4) it differs from most of the above existing approaches, and can be known as an *incremental anomaly detection approach*. Based on an exhaustive survey of published incremental anomaly detection approaches, Bhuyan *et al*. conclude that most existing approaches have a high rate of false alarms, are non-scalable, and are not fit for deployment in high-speed networks [13]. On that perspective, the proposed approach differs from existing incremental approaches, since this is scalable in terms of storage and possible to incorporate with live analysis on high-speed networks. The proposed approach requires maintaining only a single value for a given node. Most of the current intrusion detection approaches do not accommodate integrating contextual information with attack detection and attribution and are heavily dependent on technical traits only [13-20]. Hence, our approach is significantly different from most of the existing approaches. However [8-12,22-24] can be identified as deviations from the current general and quite simple systems.

Kandias *et al*. propose a model to integrate the user's technological traits with data obtained from psychometric tests [24]. Although the authors focus on insider attacks, the core idea in their paper coincides (to some extent) with our work, since they do not depend completely on network traces. They combine users' (psychological) profiles with technical data. However, their model is highly subjective, organisationally dependent and does not accommodate any information gathered from contextual analysis. Most importantly, it cannot be applied to profile non-human actors. In contrast, ours can be used to profile human, non-human or even virtual actors and can be extended to accommodate a wide range of contextual information.

Chivers *et al*. provide a scalable solution to identify suspicious slow insider activities, combining evidence from multiple sources using the well-known Bayes' formula [8,9]. Although similarly motivated, our work mainly differs from the decision criteria used for the analysis as described in Section 3 and from the target domain. Also, we have discussed the possibility of extending the same formula to integrate contextual information on detection. Chivers *et al*. distinguish between anomaly and normal behaviours by setting a control (base line) and choosing the one most deviant from the control as an attacker. This is not practical, as it is very hard setting a predefined baseline for node behaviours and the authors have not discussed it. As we identified, when there are more than one attacker in a subnet with higher variations of nodes behaviours, this decision criterion does not work well. Comparison across subnets (i.e. using a common baseline for all subnets) is also problematic. Identifying anomaly nodes through visually inspected row score graphs is another issue in Chivers *et al*.'s work. Such a decision can be affected by even dimensions of the drawing canvas in a situation where there is a higher

variation in parameter values. In such a situation, *standardisation* of node scores should be performed, before any comparison, which has been ignored by their work. However Chivers *et al*. themselves identify a need for different decision criteria other than the *Maximum score function* method they used. We have incorporated the concept of statistical normality into our work when addressing these issues.

Basu *et al*. propose an approach which uses connection-based windows to detect low-profile attacks with a confidence measure while Streilein *et al*. use multiple neural network classifiers to detect stealthy probes [22,23]. [21,24,29] can be identified as much more similar studies to Chivers *et al*.'s work. In [23,24], users are profiled according to their behaviour and that information is used to identify users who require further investigations. Evidence accumulation as a means of detecting slow activities has been proposed by [21]. All the above approaches, except [8,9,21], require the storage of large volumes of event data for later analysis, and hence differ from our work. [21] differs from our work as it uses a counting algorithm instead of the Bayesian approach and also in its decision criteria. Importantly, all the above approaches, except [24], are profiling the suspected origins based on technical solution data only. Since our aim is not only to propose an efficient attribution methodology but also to conduct an investigation of its effectiveness under different conditions, certainly this work significantly differs from all the above works.

3. METHODOLOGY

We address the problem by dividing it into two separate smaller sub-problems: *Evidence fusion & aggregation (Accumulation)* and *Analysis (Anomaly definition)* assuming that exiting signature detection algorithms could be employed to detect the events (signature elements) of an attack pattern. The term node is used in this paper to denote anything in terms of identities, which can be a user, machine, account number or a location (physical or virtual), essentially the *visibility source* of a potential attack [2,3].

A. Evidence fusion & aggregation

According to Brackney *et al*., integrating information from many sources in a manageable and scalable fashion, in order to identify patient attackers, is still an important open question [18]. Chivers *et al*. claim that combining events from one or more sensors (possibly of various types) while reducing data without adversely impacting detection is a major challenge [8,9]. Both statements are talking about 'Evidence fusion & aggregation'. Chivers et al. use a Bayesian approach, while [21] uses a counting algorithm for this purpose. However [8,9] show that the Bayesian approach is superior to the counting algorithm. At this stage, we also used the simple Bayes' formula for evidence fusion, as described in the next sub-section. Jiang et al. show that probabilistic correlation works well in noisy environments [28]. However, investigating ways to apply other possible methods, instead of the simple Bayes' formula, such as Bayesian Belief network, Kernel Density Estimation (KDE), Dempster-Shafer theorem, Kalman Filter, Viterbi algorithm, Gi*, Evidential reasoning, Logic based fusion, Preference aggregation, Neural networks, Ontology & category theory for this task would be interesting and is left as future work in this ongoing work.

Bayesian approach

The posterior probability of the hypothesis H_k given that E is given by the well known formula:

$$P(H_k/E) = \frac{p(E/H_k) \bullet p(H_k)}{p(E)} \quad (1)$$

In order to fit this formula into our case, let H_k: hypothesise that k^{th} Node is an attacker and $E = \{e_1, e_2, e_3, \ldots, e_m\}$ the set of all suspicious evidence observed against node k during time t from m different independent observation spaces. Here $P(E)$ is the probability of producing suspicious events by node k, but on its own is difficult to calculate. This can be avoided by using the law of total probability and reformatted (1) as:

$$P(H_k/E) = \frac{p(E/H_k) \bullet p(H_k)}{\Sigma_i\, p(E/H_i) \bullet p(H_i)} \quad (2)$$

For independent observations, the joint posterior probability distribution:

$$P(H_k/E) = \frac{\Pi_j\, p(e_j/H_k) \bullet p(H_k)}{\Sigma_i\, \Pi_i\, p(e_j/H_i) \bullet p(H_i)} \quad (3)$$

Once we observed E from node k, to calculate the posterior probability of node k being an attacker $p(H_k/E)$, it is necessary to estimate:

1. $p(e_j/H_i)$ - likelihood of the event ej given the hypothesis H_i and,
2. $p(H_i)$ - prior probability

Assuming that we know the prior and likelihoods, it is obvious that (3) facilitates to combine evidence from multiple sources (contextual information) to a single value (posterior probability) which describes our belief, during a short observation period, that node k is an attacker given E. Aggregating short period estimations over time helps to accumulate relatively weak evidence for long periods. This accumulated probability term, $\Sigma_t\, p(H_k/E)$ (t is time) known as *profile value* hereafter, can be used as a measurement of the level of suspicion for node k at any given time. Schultz *et al*. claim that profiling suspected insiders provides one of the best ways of reverse engineering an attacker [25]. Although there are some significant differences between the characteristics of insiders and outsiders, profiling can still be used effectively in cyber conflict attribution, as shown in the rest of the paper.

B. Analysis

At any given time, given the profiles of all nodes, detecting suspicious profiles is the analysis stage as the attacker's activity pattern is now reflected by profiles. Bhuyan *et al*. claim that anomaly detection is usually flexible and sufficient to detect both unknown (novel) and known attacks [13]. When there is an attacker who violates legitimate users' activity patterns the probability that the attacker's activity is detected as anomalous should be high. We distinguish between anomalous and normal profiles using the concept of statistical normality.

Statistical Normality

The statistical approach to *normality* defines it in terms of a normal distribution curve. A normal curve is a statistical data distribution pattern occurring in many natural processes. As long as what is most common (average or most frequent) in the general population is considered as normal, any behaviour or characteristic that occurs only rarely can be regarded as abnormal. In a normal distribution, node profiles lying outside (around) three standard deviations from the mean can be considered as abnormal. This boundary may vary, so one may define abnormality beyond two standard deviations from the mean and hence select a wider selection of nodes for further investigation. One advantage of this is that confidence in attribution can also be expressed in probability terms. Calculating standardised node profiles (Z-scores) instead of node profiles themselves, will resolve the analysis problem better.

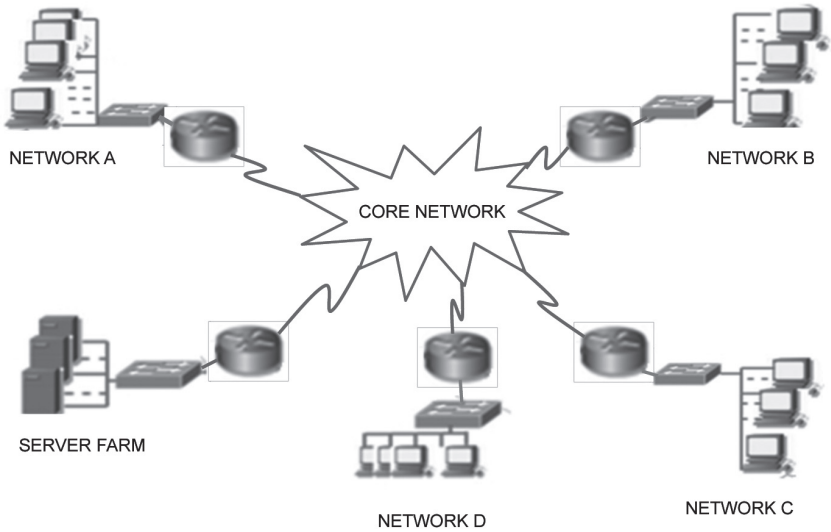
4. EXPERIMENTAL SETUP

To demonstrate the proposed approach, a series of experiments were conducted. Simulation was used to express network topology and traffic patterns of interest were generated using NS3 [26], assuming Poisson arrival model with inter-arrival time gap between two consecutive events as an exponential, to collect data on the characteristics and behaviour of several common network reconnaissance tools. Each simulation was run for a reasonable period of time to ensure that enough traffic was generated (over one million events).

A. Network Topology

Figure 1 shows the network topology used for our experiments. A total of 2,122 nodes were distributed among four networks labelled *A* (99 nodes), *B* (400 nodes), *C* (800) and *D* (800 nodes). In addition, a network dedicated to a server farm was simulated with 23 nodes.

FIGURE 1. THE NETWORK TOPOLOGY USED FOR EXPERIMENTS.
SOURCE FOR GRAPHIC SYMBOLS: FUNDAMENTALS OF NETWORK SECURITY GRAPHIC SYMBOLS, CISCO NETWORKING ACADEMY PROGRAM (FREELY AVAILABLE ON WWW).



B. Attacker Modelling

If λ_s . λ_1 are mean rates of generating suspicious events by suspicion and normal nodes respectively, we ensured maintaining $\lambda_s = (\lambda_1 . \lambda_1 + 3\sqrt{\lambda_1}]$ and $\lambda_1(=0.1)$ sufficiently smaller for all our experiments to characterise slow suspicious activities which aim atstaying beneath the threshold and hiding behind the background noise. $\sqrt{\lambda_1}$ is the standard deviation of rates of suspicious events generated by normal nodes.

C. Parameter Estimation

Prior probabilities and Likelihoods are assigned as follows.

$$P(H_m) = P(H_n) = \frac{1}{\text{Number of nodes in the scene}} \text{ , for all m, n and } m \neq n \quad (4)$$

$$p(e_j/H_m) = p(e_j/H_n) = k, \quad \text{for all j, m, n and } m \neq n \quad (5)$$

(4) assumes that all nodes in the scene have a same prior belief (equally likely) to be subverted. However, this is not the case in many situations. In cyber warfare, as many countries have a cold cyber war with other countries [6], one entity may have a higher prior belief of suspicion about the activities of another. In networks, an e-commerce server may have a higher chance to be subverted than a client node. In a company, an angry programmer attached to the IT department could be more dangerous than a loyal employee in the marketing department. Therefore if the analyst requires to distinguish between identities (or clusters of identities, for example, in case of identity is a geospatial location; a cluster can be a province, a country or even an alliance of countries), prior probability can be assigned separately. Since prior probabilities are based on previous experiences, $p(H_m)$ can be judged by the analyst, based on the information gathered from contextual analysis or intelligent services.

(5) explains the likelihood of producing event ej by any node if it is subverted. For the purpose of demonstration, we assigned arbitrary values (≤ 1) for k. However it can be estimated as follows. If ej is an event such as *UDP scan* or *land attack* which cannot be expected from a non-subverted node, then k can be assigned to one. However, k cannot always be one, for some suspicious events that appear as a part of attack signatures could also be originated from normal network activities. For example, a major router failure could generate many ICMP unreachable messages; an alert of multiple login failures could result from a forgotten password. An execution of *cmd.exe* could be part of a malicious attempt or a legitimate one, as it is frequently used by malicious programs to execute commands while it is also frequently used by legitimate users during their normal day-to-day operations. The question is how to estimate $p(e_j/H_m)$ if e_j becomes such an observation (true positives)? One possible answer would be using IDS evaluation datasets such as ISCX 2012 [32] or DARPA as corpuses and using similar techniques used in the natural language processing domain. Chivers *et al*. claim that, in some cases, the historical rate of occurrences of certain attacks is known and can be used to estimate the likelihood that certain events derive from such attacks or it may be sufficient to quantify

these frequencies, in a similar way to estimating risk likelihoods, to an accuracy of an order of magnitude [9].

5. RESULTS AND ANALYSIS

In this section, experimental results are presented along with the analysis.

A. Identifying Suspicious Nodes

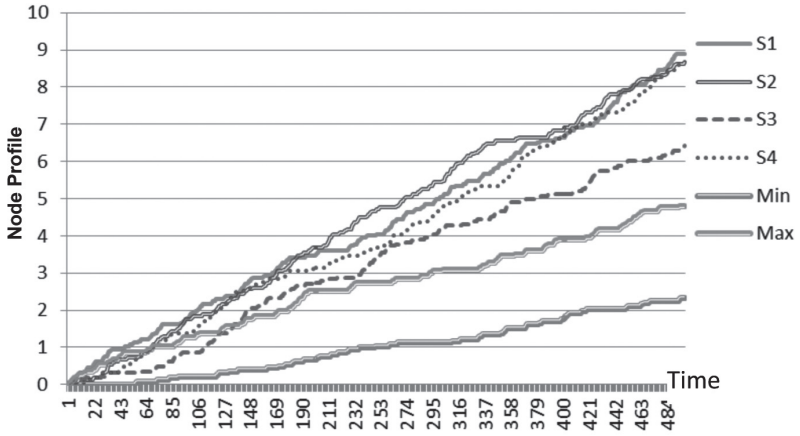
The proposed approach was tested against 25 (=5x5) test cases, varying subnet sizes {25,50,100,250,500} and number of attackers {1,2,4,7,10}, and it was observed that the proposed approach detected slow attackers well in all 25 cases. Due to space constraint only one test case, 100 size subnet with four attackers, is listed here.

Four low rate attackers were located in a 100 size subnet of network B. All clients generated innocent events (events such as forgotten password etc.) while four attackers generated low rate attack (reconnaissance) events. At each time point, node profiles were calculated for all 100 nodes in the subnet and converted to Z-scores. Node profiles and Z- scores were plotted as in Figures 2 and 3 respectively.

1) Maximum Score approach

As mentioned in Section 2, selecting suspicious nodes by looking at raw node profiles is problematic when there is more than one suspicious node. Although all suspicious nodes are above the *Max* line (after some time), setting this *Max* is problematic in real world implementations.

FIGURE 2. CUMULATIVE PROBABILITIES (NODE PROFILES), S1,S2,S3,S4 DENOTE ATTACKERS. MIN AND MAX REPRESENT THE MINIMUM AND MAXIMUM Z-SCORES OF NORMAL NODES AT EACH TIME POINT.



2) Z-Score approach

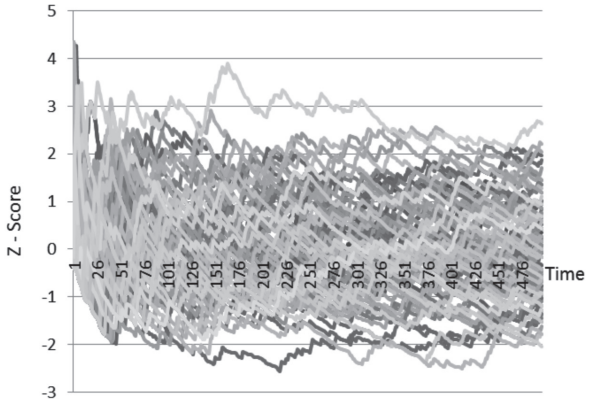
Attackers are always above or near around three standard deviations from the mean, and most importantly, there is a clear visual separation between a set of normal nodes and anomaly nodes. Graphs become more stable by the time (i.e. assuming stationary status), which means the proposed decision criteria are better for distinguishing anomalous profiles from normal profiles than the 'Maximum score approach'.

FIGURE 3. Z-SCORES OF NODE PROFILES. S1,S2,S3,S4 REPRESENT SUSPICIOUS NODES. MIN AND MAX REPRESENT THE MINIMUM AND MAXIMUM Z-SCORES OF NORMAL NODES AT EACH TIME POINT.

3) Best and worst cases

To investigate how the proposed approach works with best and worst cases, the above experiment was repeated twice, first without any attackers and then with all subverted nodes, and obtained the similar graphs as in Figure 4 in both cases. Most of the nodes are nearly between three standard deviations from the mean, and none of the nodes can be seen clearly separated from the majority. However this would not be a problem. If an analyst sees a similar graph, it would be safe to assume that all nodes are subverted (instead of assuming they are free of attackers) and to do further investigations on one or two nodes to verify. If investigated nodes are attackers it is reasonable to consider that all nodes are attackers or vice versa.

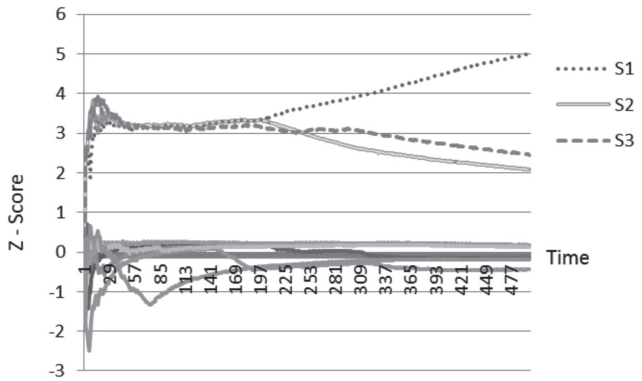
FIGURE 4. Z-SCORES OF NODE PROFILES, NO ATTACKERS, 100 SIZED SUBNET.



4) Node behaviour

To investigate how proposed Z-score graphs reflect the behaviour of nodes (identities), three attacker nodes were located in a 50 size subnet in network D. All others were innocents. Two out of three attackers stopped their reconnaissance attempts at 200 and 300 times respectively. As shown in Figure 5, when an attacker node changes its behaviour the relevant Z-score graph responds to that behaviour by changing its direction.

FIGURE 5. Z- SCORE GRAPHS ARE SENSITIVE TO NODE BEHAVIOUR. S1,S2,S3 ARE SUSPICIOUS NODES. ALL OTHERS ARE INNOCENTS.



B. Attacker Localisation

In a situation, there are multiple suspected sites to be investigated (e.g. different actors, subnets, LANs, locations etc) and determining the centres of attention would be problematic. Localisation of attackers' identities as much as possible, at least for an intermediary level, or choosing the smallest subset in which an attacker may be located, would greatly save the cost and time to be spent on investigations. To investigate the capability of the proposed approach herein: one attacker was placed in a subnet of network C. Scores were assigned (profiling) the Gateways of each subnet, using the formula:

$$\textit{Gateway score} = \frac{\textit{Cumulative Score}}{\textit{Number of nodes in the subnet}}$$

assuming each reconnaissance event can be reverse engineering only up to the gateways. They were converted to the Z-scores and Figure 6 was obtained. GA, GB, GC and GD are gateways of networks A, B, C and D respectively.

FIGURE 6. Z – SCORES OF GATEWAY SCORE OF EACH NETWOK.

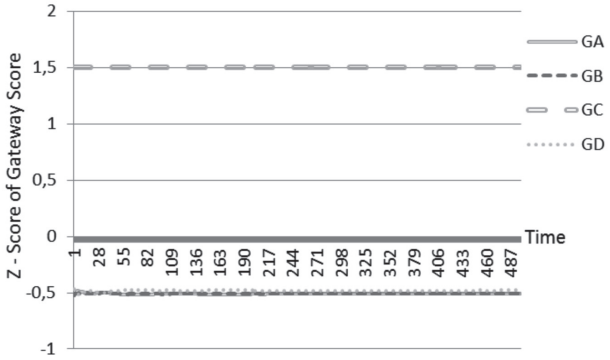


Figure 6 proves the proposed approach useful in attacker localisation.

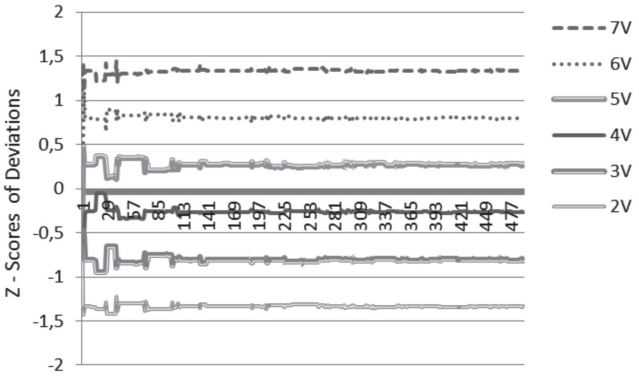
C. Network Parameters

In this section we investigate how different network parameters affect the attribution of slow activities.

1) Traffic Volume

An attacker was located in a 51 size subnet of Network C and generated events. The same experiment was repeated six times, keeping all parameters unchanged except the attacker's traffic volume. If the attacker's traffic volume is V the first time, then at each repetition the attacker's traffic volume was incremented by one time as *2V, 3V, ...,7V*. For each experimental run, the deviation of attacker's profile value from the average of normal (statistical norm) was calculated. Then the standardised deviations (z-scores of deviations) are plotted as in Figure 7. The graph tells us: 'the higher the traffic volume generated by attacker, the easier his detection will be.'

FIGURE 7. Z-SCORES OF DEVIATIONS OF CUMULATIVE NODE SCORES.



2) Cluster Size

To investigate how the identities' cluster size (here subnet size) affects detection, an attacker was located in a 500 size subnet and the same experiment was repeated six times by keeping all other parameters, except the subnet size, unchanged. Subnet size was changed to 400, 300, 200, 100, 50 and 25 at each experimental run and the graphs in Figures 8, 9 and 10 were obtained. Figure 8 and 9 say 'attackers have less chance to hide behind innocent events, when the cluster size decreases.' It is further reinforced by Figure 10 saying 'the smaller the cluster size, the better for detection of suspicious slow activities' in terms of security. But, in practice, it should be noted that partitioning a network into very small subnets would not be a feasible solution sometimes, as it depends on several other factors such as resources availability and user requirements. Figure 10 also suggests that 'going beyond 100 size cluster would not make any real sense in terms of detection.'

FIGURE 8. PERCENTAGES (%) OF SUSPICIOUS EVENTS GENERATED BY ALL INNOCENTS.

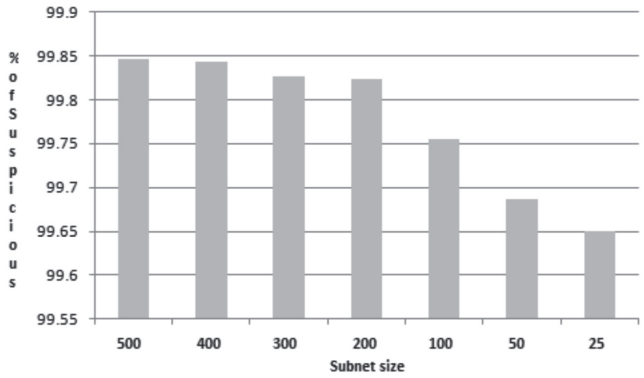
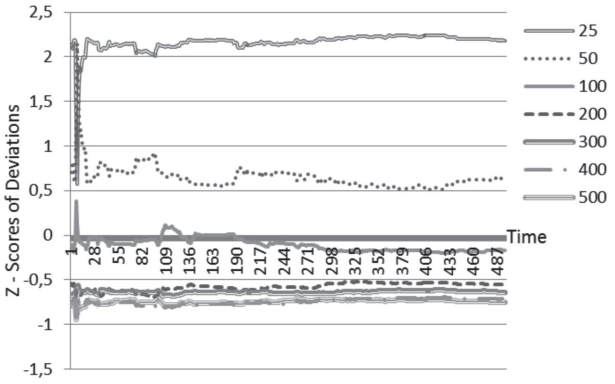
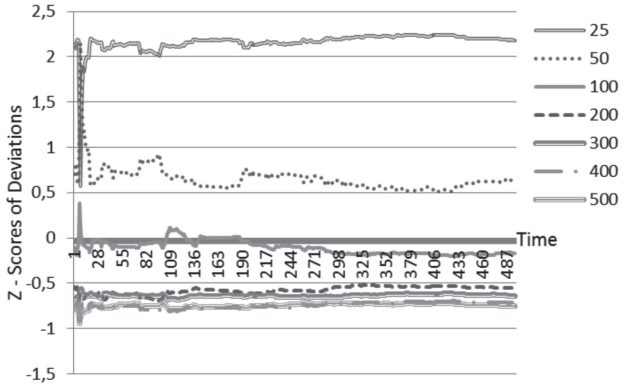


FIGURE 9. PERCENTAGES (%) OF SUSPICIOUS EVENTS GENERATED BY ATTACKER.



The authors would like to reiterate that a subnet equals a cluster of identities. For example, in a case of cold cyber war or in an attack like the well-known Georgia 2008 case, a cluster can be a country or a region of a suspected country and identity can be any physical or virtual location within that country or region.

FIGURE 10. Z – SCORES OF ATTACKER'S DEVIATIONS FROM THE AVERAGE.



3) Number of Attackers

Keeping all conditions unchanged, except number of attackers, the same experiment was repeated twice, first with two attackers and then with seven attackers. The attacker's node score (see Figures 11 and 12) is dependent on 'the number of attackers in his own subnet' (compare attackers' Z-scores). This rationalises the usage of 'Statistical normality' as the decision criteria and suggests defining 'one's abnormality' relative to his peers (i.e within the same domain, department, similar user group, region, country etc.) would give better results (in terms of lower false alarms) than defining it universally. Comparison of nodes profiles (as in Figure 2) regardless of their subnets would give higher false alarms.

FIGURE 11. Z-SCORE GRAPHS FOR SAME SIZE SUBNETS WITH DIFFERENT NUMBER OF ATTACKERS (250 SIZE SUBNET, TWO ATTACKERS)

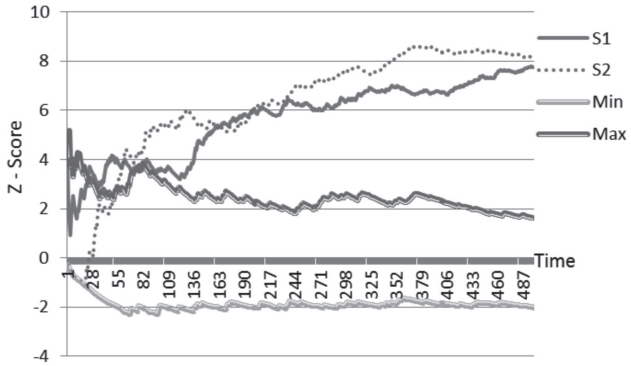
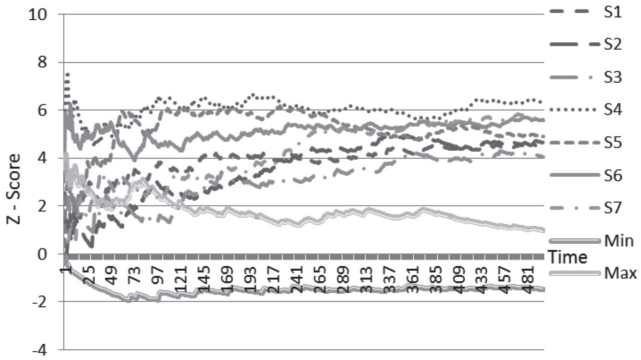


FIGURE 12. Z-SCORE GRAPHS FOR SAME SIZE SUBNETS WITH DIFFERENT NUMBER OF ATTACKERS (250 SIZE SUBNET, SEVEN ATTACKERS).



6. SAMPLING TECHNIQUES

Many IDSs such as Snort facilitate for logging data in a variety of ways for later analysis, as it is an essential part of any intrusion detection activity. If you are not looking at the logs and monitoring the alerts, then effort invested into an IDS can quickly become meaningless [27]. In a slow attack environment, logging is crucial as you cannot log everything during longer times. The large size/unmanageable nature of the target population is one of the main reasons for sampling instead of doing a census. As it is almost similar to the problem the analyst faces herein, the simple random sampling technique was used to investigate the usability of sampling for data logging in slow-attack environments.

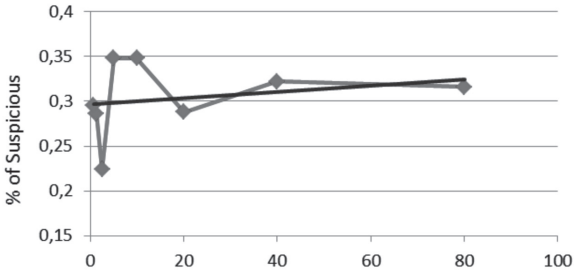
An attacker was located in a subnet of network C and 'stateless' attacks events were generated. The simulation was allowed to run 1440 time units. The whole period was divided into twelve blocks, and within each block, a sample was collected using an R [31] script. Finally, all twelve samples were combined together to make one final sample. The same experiment was repeated with different sample sizes in order to identify how sample sizes affect 'detection potential.' Table 1 and Graphs in Figures 13, 14, 15 and 16 show the experimental outcomes. We varied the sample sizes from 80% to 0.625% (see Table 1), always half of the previous size.

TABLE I. SAMPLING STATISTICS.

Sample Size as a % of population/whole observation)	80	40	20	10	5	2.5	1.25	0.625
Number of Attack Events selected	826	420	188	113	56	18	12	6
Number of Innocent Events selected	260244	130235	65200	32356	16043	8029	4188	2026
Percentage (%) of Attack Events	0.32	0.32	0.29	0.35	0.35	0.22	0.29	0.30

Although the fitted trend line in Figure 13 shows a very small positive trend between percentage of suspicious events and sample size, the real figures in the table explain that it would not be significant. Interestingly, 'in each sample, the percentage of suspicious events generated by the attacker is almost same as it is in the population (0.3)' is a good indicator that selected samples represent the intended population's characteristics, regardless of its size. Analyst may choose sampling techniques for long-term networking monitoring (it could not be for detection, but may be for other purpose of traffic analysis), deciding the sample size based on the resources availability and the intended purpose.

FIGURE 13. PERCENTAGE OF SUSPICIOUS EVENTS GENERATED BY ATTACKER.



Graphs in Figure 14, 15, 16 show that the analyst can enjoy the population characteristics (in terms of this analysis) even if the size of the sample is 5% of the entire data capture. This would be a good indicator, why?, if an analyst can reduce his focus by 95% it will reduce the time and cost too. However when the sample size is smaller than 2.5% of its population size, anomaly-based detection methods cannot be used. But the table explains that signature based detection methods can still be used, as it contains very few attackers' signatures. Generally using 10% size sample would be an ideal for detecting suspicious slow activities, whether it is based on anomaly or signature-based detection methods. However the authors do not generalise the optimal sample size as 10%. It could be highly subjective and varied according to the intended analysis. Further experiments are needed on this topic. At least at this stage, the authors have shown that some population characteristics remain unchanged in samples and, hence there is a possibility to use sampling techniques in this domain.

FIGURE 14. Z-SCORES, WHEN THE SAMPLE SIZE IS 10% OF WHOLE TRACE. S REPRESENTS THE SUSPICIOUS NODE. MIN AND MAX REPRESENT THE MINIMUM AND MAXIMUM Z-SCORES OF NORMAL NODES AT EACH TIME POINT.

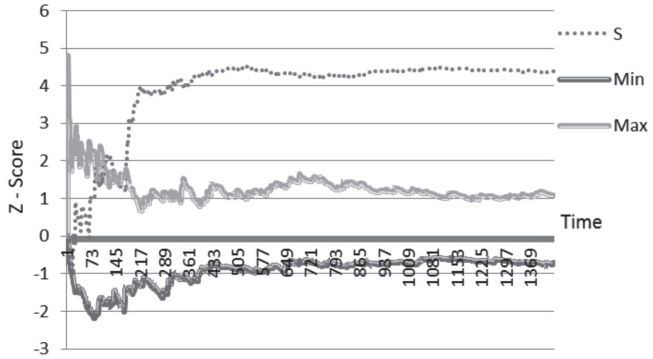


FIGURE 15. Z-SCORES, WHEN THE SAMPLE SIZE IS 5% OF WHOLE TRACE. S REPRESENTS THE SUSPICIOUS NODE. MIN AND MAX REPRESENT THE MINIMUM AND MAXIMUM Z-SCORES OF NORMAL NODES AT EACH TIME POINT.

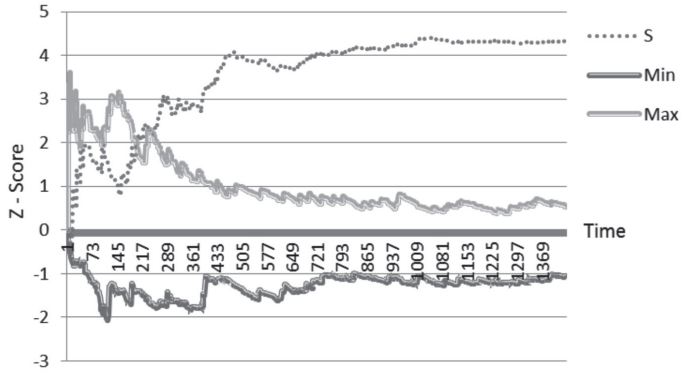
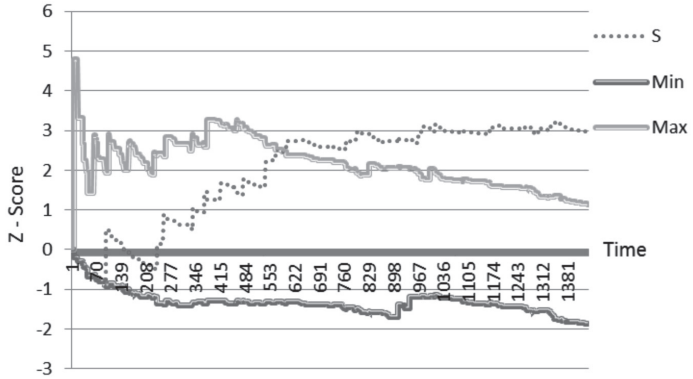


FIGURE 16. Z-SCORES, WHEN THE SAMPLE SIZE IS 2.5% OF WHOLE TRACE. S REPRESENTS THE SUSPICIOUS NODE. MIN AND MAX REPRESENT THE MINIMUM AND MAXIMUM Z-SCORES OF NORMAL NODES AT EACH TIME POINT.

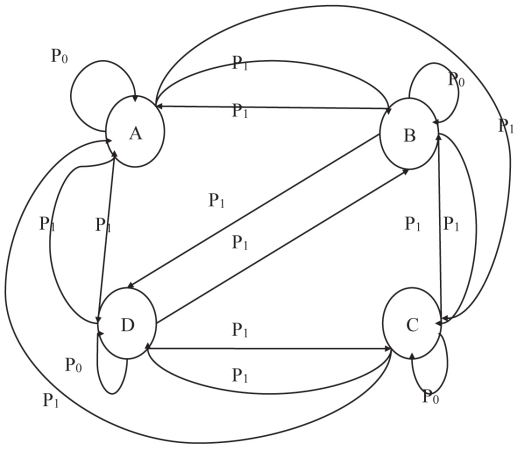


7. DISCUSSION

An efficient method for cyber conflict attribution (particularly slow activities) and an investigation of its effectiveness under different conditions have been provided. Breaking down the attribution problem into two sub-problems reduces the complexity of the problem, and explores ways to investigate alternative methods. The proposed approach is domain agnostic. It can be easily adjusted to use in many aspects of cyber warfare and help in actor intelligence: profiling adversarial technical capabilities; creating linkage between actor groups; tracking the supply chain; and differentiating between actors (e.g. state-sponsored or criminal) etc. It can be used for profiling any kind of actors, not only in the cyber domain but also in other domains such as crime and juridical sciences. Experimental outcomes and recommendations presented in Sections 5 and 6 provide tactical and operational principles for systematic and

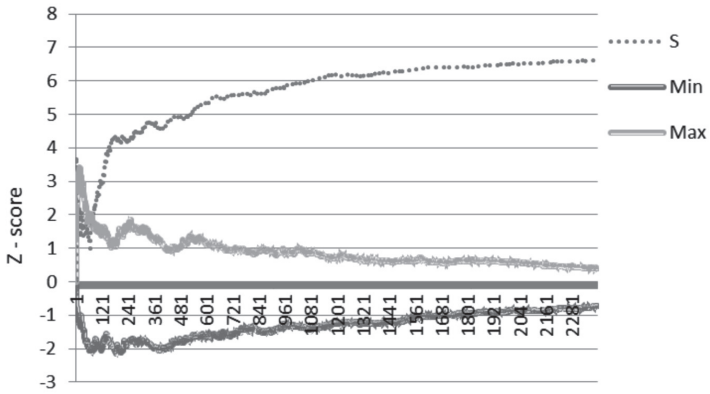
efficient profiling and attribution. They are particularly useful in the capacity planning stage of a network design process. Findings of how cluster size affects detection can be incorporated with existing clustering based analysis approaches [30,14]. In future, identifying the best performance method (among alternative methods such as using sensor fusion algorithms) and handling some miscellaneous issues, such as overcoming situations when the source of the event is unknown, will be addressed. Based on the idea derived from Section 6, an experiment was set up to investigate the possibility of using mobile sensors to slow activity detection. An attacker was located in network D. A Finite state automaton (see Figure 17) was used to control the sensor mobility (transitions). At any given state, the sensor spends a constant time interval for monitoring. Scores were updated only when the sensor had visibility to the target subnet.

FIGURE 17. FINITE STATE AUTOMATA USED FOR SENSOR MOBILITY, P_0=0 AND P_1=0.33.



As Figure 18 shows, it can identify the attacker, even using a mobile sensor. This could be mainly due to the cumulative nature of the proposed approach and the usage of automaton. It should be noted that the transition probabilities *(P_0, P_1)* of the above automaton can be estimated dynamically, based on evidence at the scene, in order to improve the quality of the detection, which is also left for future work.

FIGURE 18. Z-SCORE GRAPH FOR SENSOR MOBILITY. S REPRESENTS THE SUSPICIOUS NODE. MIN AND MAX REPRESENT THE MINIMUM AND MAXIMUM Z-SCORES OF NORMAL NODES AT EACH TIME POINT.



REFERENCES:

[1] S.W. Beidleman, 'Defining and Deterring Cyber War,' Msc Thesis, Dept. Military Strategy Planning and Operations, U.S. Army War College, Carlisle, Pa, 2009.

[2] D.Morrill. (2006, August 07). Cyber Conflict Attribution and the Law [Online]. Available: http://it.toolbox.com/blogs/managing-infosec/cyber-conflict-attribution-and-the-law-10949

[3] D.A.Wheeler and G.N.Larsen. (2003, October 30). Techniques for Cyber Attack Attribution. Inst. for Defense Analyses [Online]. Available: http://www.dtic.mil

[4] T.Parker. (2010). Finger Pointing for Fun, Profit and War? The importance of a technical attribution capability in an interconnected world [Online]. Available FTP: media.blackhat.com Directory: /bh-dc-11/ Parker File: BlackHat_DC_2011_Parker_Finger_Pointing-wp.pdf

[5] S.Charney. (2009). Rethinking the Cyber Threat: A Framework and Path Forward [Online]. Available FTP: download.microsoft.com Directory: download File: rethinking-cyber-threat.pdf

[6] K.Saalbach. (2011). Cyberwar Methods and Practice [Online]. Available FTP: dirk-koentopp.com Directory: download File: saalbach-cyberwar-methods-and-practice.pdf

[7] N.Villeneuve and D.Sancho. (2011, September 26). The Lurid Downloader [Online] Available: http://www.trendmicro.com

[8] H.Chivers et al., 'Accumulating evidence of insider attacks,' in The 1st International Workshop on Managing Insider Security Threats 2009 (In conjunction with IFIPTM 2009) CEUR Workshop Proc., 2009, pp.34-51.

[9] H.Chivers et al., 'Knowing who to watch: Identifying attackers whose actions are hidden within false alarms and background noise,' Inform. Syst. Frontiers, Springer, 2010, doi: 10.1007/s10796-010-9268-7.

[10] A.Patcha and J.M.Park, 'An overview of anomaly detection techniques: Existing solutions and latest technological trends,' Elsevier Computer Networks, Vol. 51 Issue 12, pp. 3448–3470, August 2007.

[11] S.Kumar and E.H.Spafford, 'An application of pattern matching in intrusion detection,' The COAST Project, Dept. Comp. Sci, Purdue University, Tech. Rep, 1994.

[12] V.Chandola et al., 'Anomaly detection: A survey,' ACM Computing Surveys (CSUR), Vol. 41 Issue 3, July 2009, doi:10.1145/1541880.1541882.

[13] M.H.Bhuyan et al., 'Survey on Incremental Approaches for Network Anomaly Detection,' IJCNIS Vol. 3, December 2011, pp 226-239.

[14] C. Zhong and N. Li, 'Incremental Clustering Algorithm for Intrusion Detection Using Clonal Selection,' in Proc. PACIIA (1), 2008, pp.326-331.

[15] F.Ren et al., 'Using density-based incremental clustering for anomaly detection,' In: CSSE '08, DC, USA, IEEE Comput. Soc., 2008, pp. 986-989.

[16] R.Bejtlich, The Tao of Network Security Monitoring: Beyond Intrusion Detection, Addison-Wesley, 2005.
[17] A.J. Beecroft, 'Passive Fingerprinting of Comput. Network Reconnaissance Tools,' MSc Thesis, Naval Postgraduate School, Monterey, California, 2009.
[18] R.C.Brackney and R.H.Anderson, 'Understanding the insider threat,' Proc. March 2004 Workshop, RAND Nat. Security Research Division,Tech. Rep., 2004.
[19] W.Y.Yu and H.M.Lee, 'An incremental-learning method for supervised anomaly detection by cascading service classifier and its decision tree methods,' in PAISI '09 Proc. Pacific Asia Workshop on Intell. and Security Informatics, 2009@Springer Berlin /Verlag, doi:10.1007/978-3-642-01393-5_17.
[20] P.Laskov et al., 'Incremental support vector learning: Anal., implementation and applications,' J. of Machine Learning Research Vol.7, October 2006, pp. 1909-1936.
[21] T.Heberlein, 'Tactical operations and strategic intelligence: Sensor purpose and placement,' Net Squared Inc, Tech. Rep. TR-2002-04.02, 2002.
[22] W.W.Streilein et al., 'Improved detection of low-profile probe and novel denial-of-service attacks,' Int. Workshop on Statistical and Machine Learning Techniques in Comput. Intrusion Detection, 2002.
[23] R.Basu et al., 'Detecting low-profile probes and novel denial-of-service attacks,' IEEE SMC IAS Workshop, West Point, New York, USA, Tech. Rep., 2001.
[24] M.Kandias et al., 'Dimitris gritzalis: An insider threat prediction model,' in Trust, Privacy and Security in Digital Business, 2010@Springer Berlin/ Heidelberg, doi: 10.1007/978-3-642-15152-1_3.
[25] E.E.Schultz and R.Shumway, Incident response: A strategic guide for system and network security breaches, Indianapolis: New Riders, 2001.
[26] The NS3 discrete-event network simulator [Online]. Available: http://www.nsnam.org/
[27] J.Babbin et al., Snort Cookbook [Online]. Available: http://commons.oreilly.com/wiki/index.php/Snort_Cookbook
[28] G.Jiang and G.Cybenko, 'Temporal and spatial distributed event correlation for network security', Proc. Amer. Control Conference, Boston, MA, 2004, pp. 996-1001.
[29] P.G.Bradford et al., 'Towards proactive computer system forensics', Int. Conference on Information Technology: Coding and Computing, IEEE Comput. Soc., 2004, pp.648-653.
[30] M.C. Libicki (2009).Cyberdeterrence and Cyberwar. Santa Monica, CA: RAND Corporation [Online]. Available: http://www.rand.org/pubs/monographs/MG877
[31] The R project for statistical computing: R foundation for statistical computing [Online]. Available: http://www.r-project.org
[32] A. Shiravi et al.. (2012). ISCX Intrusion Detection Evaluation DataSet [Online]. Available: http://www.iscx.ca/datasets

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

The Role of COTS Products for High Security Systems

Robert Koch
Institut für Technische Informatik
Universität der Bundeswehr
Neubiberg, Germany
Robert.Koch@UniBw.de

Gabi Dreo Rodosek
Institut für Technische Informatik
Universität der Bundeswehr
Neubiberg, Germany
Gabi.Dreo@UniBw.de

Abstract: Today, economic pressures and decreasing military budgets enforce a revision of armament projects. While comprehensive and cost-intense equipment acquisitions could be realised during the Cold War, overextension and global economic crisis forced the abatement of projects and broad cutbacks of the residual undertakings.
Based on this, one of the most important tendencies of the last few years is the intense use of commercial off-the-shelf products (COTS) and master agreements with the industry. In contrast to past armament projects, the necessary hardware and software is no longer designed with respect to special military requirements, but products already available on the market are used wherever applicable.
The increasing use of COTS products in all areas of armament is a matter of special importance, opening tenuous points of attack. By that, the number of important security incidents has grown larger in the past few years even with more and improved security mechanisms like firewalls and Intrusion Prevention Systems in place.
On the contrary, through the use of sophisticated and targeted attacks, even highly secured or isolated networks and systems can be compromised. Stuxnet or the attacks on RSA and the subsequent compromise of Lockheed Martin and other companies of the American defence industry are well-known examples. Conficker was another demonstration of the comprehensive infection of secured networks, for example in the Federal Armed Forces or the Royal Navy.
Based on that, a significant security hazard arises which is of essential importance with regard to the Cyber Domain.
This paper analyses the effect of COTS products and proprietary software with respect to the security of military systems. Based on the identified endangerments, conclusions for the recovery of the security of military information systems are presented and implications for the implementation of Cyber Operations are given.

Keywords: *commercial off-the-shelf, high security, data leakage, information operations*

1. INTRODUCTION

Armament projects are often characterised by their complexity, typically in conjunction with high costs for development and acquirement, but also for maintenance during the utilisation period. While comprehensive and cost-intensive equipment acquisitions could be realised during the Cold War, overextension and the global economic crisis forced the abatement of numerous projects and broad cutbacks of the residual undertakings. To be able to reduce costs on new and indispensable projects, the use of COTS products and master agreements with the industry has been widely used in recent years. Therefore, hardware components are no longer designed and optimised for military applications but standard hardware products of the market are used wherever possible. On the one hand, this approach enables substantial cost savings, on the other hand, numerous problems arise which can often only be recognised at a second glance. Through the use of COTS products and widespread proprietary software, various security and supply problems are opened up which can endanger the security and availability of systems. For example, COTS products can be introduced very quickly without the need of additional development costs, but on the other side, there is often no availability guarantee.

Also, numerous systems are using proprietary software, often based on general licences concluded for whole organisational areas and running out-of-date versions of operating systems and applications. Because of the widespread use of these products in civil everyday life, these systems are alluring criminals and numerous malicious programmes are available to attack them.

The paper analyses the role of COTS products and proprietary software for high security and classified systems with respect to information security (defensive) and information operations (offensive). While there was a three-day symposium of NATO in Brussels in the year 2000 which dealt with COTS products in defence applications [1], the main focus was limited to the use of software products. However, numerous important aspects have arisen in the past few years and now, especially, the hardware has to be taken into consideration, too. Therefore, characteristic properties of COTS products – hardware and software – are presented and their vulnerabilities are analysed. After an assessment of the current situation, action needs for ensuring the security of the systems and implications for the implementation of information operations are given.

The remainder of the paper is organised as followed: First, requirements for high security networks and systems are collected in order to scale for the investigation of the role of COTS products. After that, the characteristic properties of traditional, custom-made systems as well as the aspects of COTS products and general licences are briefly described. Following, an analysis of the relevant aspects arising from the use of COTS products in high security domains and cyber operations is given. Based on these results, necessary steps for the current systems in use are drawn and conclusions for information operations are given.

2. HIGH SECURITY SYSTEMS AND NETWORKS

For implementing secure and robust systems and networks, numerous aspects of the organisational and technical domain must be considered, e.g. in the areas of management, infrastructure, systems and networks (for example, see [2]).

Several guides and recommendations can be used as a guideline to set up secure systems and networks, e.g. NIST-SP 800-36 "Guide to Selecting Information Technology Security Products" [3] or the NIST-SP 800-23 "Guide to General Server Security" [4]. From the software point of view, the basis for a secure system can be a certified Operating System (OS). For the evaluation of the security, the Common Criteria (CC) for Information Technology Security Evaluation[5] (ISO/IEC 15408) can be used. After the completion of the security evaluation, an Evaluation Assurance Level (EAL) can be achieved, where EAL1 is the lowest (functionally tested) and EAL7 is the highest (formally verified design and tested) security level. For example, the system XTS-400 Version 6.4.U4 [6] is EAL5+ certified. seL4 [7] has made a formal verification of what constitutes the basis for a certification for EAL7. Based on a secure OS, the selection of the installed programmes should be minimal and preferably also certified. A minimal set of services, protocols and software should be used.

Especially in a high-secure environment with a strict set of allowed services, the possible links between systems and servers can be monitored and controlled reliably. The use of monitoring software, anti-virus software and Intrusion Detection and Prevention Systems (IDS/IPS) is a crucial point for the surveillance of networks. Often, high security systems are isolated from other networks, or special devices like data diodes are used to secure them. However, the attacks on SCADA networks, e.g. by Stuxnet, demonstrated that offline systems and isolated networks are still not immune from attack. Therefore, the use of IDSs/IPSs is mandatory also for all kinds of critical systems. In particular anomaly-based systems can be of great use: while these systems typically suffer from high false alarm rates when used in networks connected to the Internet, these false alarms can be greatly reduced in high security networks because of the limited set of allowed services and the relatively similar communication processes. Therefore, the main reason for false alarms in traditional networks (the presence of new and unknown benign behaviour), can be excluded.

Based on the level of needed security, further requirements, for example the use of Tempest-proof hardware, high-quality cables with special characteristics regarding physical shielding or Electro-Magnetic Interference (EMI) filters can be necessary. Tempest (discovered by van Eck in 1985, therefore, also called van Eck phreaking) is the endangerment of systems because of their electromagnetic emanation which can be picked up and evaluated, compromising the data processed in a system [8]. All kinds of hardware are at risk, e.g. displays (CRT as well as LCD) [9] and keyboards [10]. Also, data cables of disk drives, etc. can be used for tapping. By using techniques like SVMs, high detection results can be achieved [11].

3. TECHNICAL ASPECTS

An important aspect in context with the origin of high costs of designed hardware is not only the development process itself or the low number of manufactured copies, but also the guaranteed availability of spare parts for a specific period in time. Therefore, the manufacturer is forced to keep spare parts or to be able to rebuild specific parts after plenty of years. Because of the long utilisation period of military equipment of about 10 to 20 years, or even longer, this can be a crucial point when using COTS: Problems can arise if spare parts are no longer available because of the short life cycles, especially in the area of the computer industry. Also, a key design goal of SCSI *is* the backwards compatibility. Therefore, an Ultra-160 SCSI disc should be usable on the bus of a quite old SCSI-1 host adapter. Even though this is possible in theory, device compatibility is often reduced in practice, for example because of different types of signalling (e.g. high and low voltage differentials). Considering other areas, these problems can grow quickly, e.g. see the development of bus architectures in PCs like ISA, VESA Local Bus, PCI, AGP, PCI-X and PCIe and their different revisions and (in)compatibilities. Therefore, it can be difficult to find specific spare parts after several years.

On the other side, the stockpiling of affected material can also be insufficient because of electrostatic sensitivity. It cannot be guaranteed that these parts are still functional after a long time of storage because of different effects, e.g. the behaviour of capacitors. Capacitors are passive electrical components, which are used to store energy in an electric field. They are used to smooth voltages on printed circuits and power supplies, etc. Typically, they consist of two conductive plates, separated by a dielectric. Often, electrolytic capacitors are used which permanently have a low loss rate. If these components are stored, the loss rate is increased based on chemical processes, e.g. the electrolyte can dry out and the capability of smoothing voltages can be reduced. Therefore, the initial current can be so high that the circuit will be destroyed when powering on the system after a few months. This effect *can* but *must not* appear. The quality of aluminium electrolytic capacitors strongly depends on the manufacturing process. The residual current behaviour is an important quantity for the recommissioning of a capacitor after an intercalation. After creating a direct current, it will be quite high and will subsequently drop down to the remaining operational power. However, by switching on the equipment, the current made can be so high that the capacitors are destroyed because of the reduced isolation capability of the dielectric and, therefore, the high leakage current. If high quality components are used, for example high-grade aluminium electrolytic capacitors, the storage time can be up to 10 years or even higher; but if only low-quality items are built in, these effects can occur even after just a few months.

For example, the impact of quality on the duration of life was analysed by a long-term study by Storelab, examining the life-time of hard disc drives (HDDs). For example, the identified failure rate of HDDs produced by Seagate was about 56 percent, while that of Hitachi was as low as five percent. Also, while the operating time of HDDs of Hitachi was about five years on average, that of Seagate drives was only 1.5 to three years, strongly depending on the specific HDD series [12].

Another aspect is the prohibition of the use of brazing solder in the European Union [13].

From July 2006, new electrical and electronic equipment must not contain lead, mercury and some other materials. For servers and storage systems, an exemption was granted until 2010; for network infrastructure equipment, e.g. switches, an exemption is still given. However, the need for using other (lead-free) materials for solder in the area of servers can have extensive consequences because the lifespan of the solder joints will be greatly reduced if they are not executed perfectly. For example, the Xbox 360 has had hardware failures in up to 50 percent of all sold units in 2006 based on problems with the lead-free solid used. This *must not* happen if high-quality components are used; however, *because* of the financial pressure and, therefore, the use of COTS, often cheap products are bought and integrated without an investigation of the installed components. Therefore, spare parts purchased at the date of the introduction of a new system can already be defective at the time of installation if they are stored for a long period. Also, inadequate air conditioning and storage can additionally reduce the lifetime of the spare parts.

Another endangerment is the used COTS hardware itself, because design and fabrication of Integrated Circuits (ICs) are typically performed by different companies to reduce costs of the fabrication process. Often only limited or no control of the manufacturing process is possible and a modification of the original design is possible. One cannot say if the specifications of a circuit contain all implemented functions or if the manufacturer retains some information. A trivial example is an Athlon-XP processor built by AMD, where a hacker found four undocumented Machine State Registers in 2010 which only could be read out after setting the Extended Destination Index to a specific value and which can be used for debugging purposes, etc. [14]. After a request, AMD confirmed the existence of undocumented registers, however, they emphasised that this is *common practice* for hardware testing and development. While no security vulnerabilities have been opened up by these registers, this example demonstrates the possibility of hidden hardware functions. To overcome this shortcoming, Bloom et al. proposed an approach to increase the trust in IC fabrication by logging forensic information of the fabrication process and printing the information on the chips, therefore, enabling an examination of deviations of the chip from the original design [15]. However, the implementation of the proposed systems requires a comprehensive adaption of the complete IC supply chain and manufacturing process for the integration of the use of a Trusted Platform Module (TPM) and corresponding runtime software. Also, several issues are not covered by the proposed approach, e.g. an insertion of trojan circuits cannot be detected, which is crucial when trying to verify the correct system behaviour of COTS in high-security systems.

Another aspect is the endangerment by pre-installed backdoors or data leakage which can be hard to detect. By the use of covert channels or techniques like steganography, an outward transfer of data can be realised which is able to easily bypass security systems. Not only can the Central Processing Unit (CPU) be manipulated in this way, but also components like the network interface card (NIC): For example, 3Com published the 3CR990 series in 2001 (after being taken over by Hewlett Packard in 2010, renamed to HP Secure), which integrates firewall functionalities directly onto the NIC. This could be a predestined point to intervene into the communication and leak data, almost impossible to detect by the server itself and only detectable by a comprehensive statistical analysis of the network traffic. For example, the timing of events can be perturbed to covertly transmit data [16], or covert channels can be encoded

directly by network packet delays [17]. An overview of covert channels and corresponding countermeasures is given in [18].

In particular, the consideration of the hardware is crucial, because the use of an EAL7-certified system is performed ad absurdum if the underlying hardware cannot be tested. It must be remembered that special security elements like a TPM chip are subject to the same problem, too, and that the correct behaviour has to be verified for the *whole* system, which is almost impossible in hindsight.

Further aspects are the software and algorithms in use. Considering government or large company projects, often master agreements are concluded – typically with market leaders of proprietary (COTS) software. On the other side, the open-source market offers a comprehensive collection of all kinds of software and algorithms. Here, two philosophies face each other: security gained by keeping an algorithm, programme, etc. secret and not giving any information about its functionality vs. opening the underlying algorithms and techniques for public examination and discussion. While the former is also known as "Security by Obscurity" and endorsed by some public institutions and industrial companies, the latter one is typically supported by scientists. Presenting an algorithm to research enables the possibility of identifying weaknesses of the design, etc. Various examples over the past few years have demonstrated that the secrecy of algorithms cannot be ensured permanently and that uncovering erroneous designs can have serious consequences, e.g. as seen by the reverse-engineering of the Crypto-1 algorithm of the Mifare-Classic RFID tags [19]. Even when Security by Obscurity can be used to temporarily disguise some limited information, like details about the infrastructure [20], using open-source and the scientific power of the community is a more promising way to gain security, as demonstrated by Hoepman et al. [21].

The correctness of the software is crucial in high-security systems. Often, a valuation of software based on the number of errors per Line of Code (LoC) is done. There are numerous arguments about which kind of software has respectively fewer programming errors, free and open source software (FOSS) or COTS. However, one always has to take into account the methodologies of the different evaluations and comparisons. For example, often only the sum of the known errors is matched, regardless of the severity of the corresponding vulnerabilities or other important aspects. For example, by investigating the details of the Common Vulnerabilities and Exposures (CVE) database [22], 48923 entries could be found on January 31th, 2012. Therefore, from 2009 to 2012, 185 vulnerabilities were identified in Windows 7, of which 47 percent can be used to gain privileges [23]. Reckoning the vulnerabilities of GNU/Linux, 429 CVEs are known from 1999 to 2012 of which 9.6 percent can be used to gain privileges. With respect to the average vulnerabilities per annum (without the CVEs of 2012 because of the early point in time of the year), Windows 7 has about 60 and the GNU/Linux about 33 vulnerabilities annually. If one considers only the vulnerabilities in GNU/Linux since the release of Windows 7 (July 2009), the average number drops down to about five. It must be taken into consideration that the statistics concerning the number of vulnerabilities often differ and the concrete numbers must be analysed in detail. For example, another evaluation mentions 299 vulnerabilities in GNU/Linux from 2009 to 2011, therefore, about 100 per year. These strong differences can arise because of the considered drivers included, e.g. most of the vulnerabilities do not originate

from the core Kernel, but from drivers of peripheries, etc. Sometimes, even utility programmes are included into the statistics, raising the numbers additionally.

Also, the severity of the vulnerabilities must be taken into consideration: For example, the possibility of gaining privileges often *can* be more dangerous than the susceptibility to a Denial of Service (DoS) attack. Therefore, the different vulnerabilities are weighted in the Common Vulnerability Scoring System (CVSS) [24] scores based on three groups (base, temporal, environmental), and their hazardousness from zero to ten with higher values presenting more serious gaps. Regarding Windows 7, the average CVSS score is 8.4, while GNU/Linux has an average about 5.3; looking forward to all vulnerabilities in the CVE database, the average is 6.9. Of course, the real-world endangerment of a vulnerability must be assessed based also on the specific requirements of the operational environment. For example, a DoS vulnerability can be more dangerous in a real-time control system than in a database system.

Also, the number of patches is sometimes used for a comparison. This is quite insufficient, because today patches are often fixing numerous security weaknesses at once, for example on fixed release circles (patch days), therefore, not opening up a comparable base.
It must at least be kept in mind that in the case of COTS software, only the released vulnerabilities can be consulted while the error search is more complex than in the case of FOSS with an available source code. Furthermore, FOSS enables numerous possibilities for security evaluation and hardening, e.g. see Charpentier et al. [25].

However, independent from the kind of software or systems in use, human beings will always produce errors. Panko gives a comprehensive overview of studies investigating how often human errors occur. In the section about programming errors, various studies are given, for example the error rate depending on the number of people in a development team or the influence of the used programming language [26]. Table 1 gives a few examples of the examined error rates.

TABLE 1: SELECTED ERROR RATES IN PROGRAMMING [26].

Reference	System / Language	Error Rate
Graden & Horsley [1986]	Major telecommunications project at AT&T, 2.5 million LoC, 8 software releases	3.7%
Linger [1994]	Formal Development / Cleanroom	0.23%
Jones [1998]	Errors per 100 LoC • Visual Basic • Java • COBOL • FORTRAN • C Average	 1.1% 1.2% 1.4% 1.6% 2.0% 1.5%
Cohen [2006]	300 code inspections CISCO systems	3.2%

Techniques like formal development and cleanroom development, etc. can help to reduce the error rates.

As virtualisation is used regularly today, it also has to be considered. On the one side, security can be enhanced by the use of virtualisation because of the isolation of different instances of OSs and applications. On the other side, the code of the Virtual Machine (VM) can also be erroneous, therefore, opening up serious vulnerabilities which can affect all running VMs. Even if there are no errors in the implementation, virtualisation concepts can be used to control systems, and are practically undetectable. The Blue Pill concept described by Rutkowska [27] is a well-known example of this kind of endangerment. Another threat that is difficult to detect derives from the use of System Management Mode-based rootkits which are able to hide their memory footprint and which are OS-independent [28].

Another important aspect is that the software can also be used to integrate backdoors – with much less effort compared with hardware. Especially when proprietary software is used and no control of the source code is possible, the risk of data leakage and pre-installed backdoors is high. The integration of rootkit-technology in DRM software on music CDs manufactured by SONY-BMG [29], or the Energizer DUO USB Battery Charger trojan which opens a backdoor on a TCP port 7777 [30] are well-known examples. Other examples can be found in the area of smartphones, where several incidents have been known in recent times, e.g. the government spying tools built into Nokia, Blackberry and iPhone smartphones as the hacking group Lords of Dharamraja released early in 2012 [31], or the rootkit software developed by CarrierIQ which is installed on approximately 140 million Android, BlackBerry and Nokia devices and acts like a spyware, e.g. logging keystrokes [32].

4. ORGANISATIONAL ASPECTS

Several organisational aspects must be taken into consideration when dealing with COTS in high-security environments. On the hardware end, by using COTS in security-sensitive systems, an important threat is opened up: because of their application area, COTS typically are not optimised or checked for radiant emittance further than the requirements of electromagnetic compatibility necessary to fulfil the directives of, e.g. the European Union transposed national laws (directive 1999/5/EC on radio equipment and telecommunications terminal equipment and the mutual recognition of their conformity [33] or directive 2004/108/EC on the approximation of the laws of the Member States relating to electromagnetic compatibility [34]). With respect to security-related systems, these directives are not sufficient: for example, electromagnetic compatibility is defined in Article 2 in 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. With respect to Annex I, "1. Protection requirements, number a, equipment shall be designed and manufactured that the electromagnetic disturbance generated does not exceed the level above which radio and telecommunications equipment or other equipment cannot operate as intended." In particular, no threshold values are given by the directives. Therefore, protection against the tapping of COTS cannot be ensured by the certified electromagnetic compatibility based on the directives. Beyond these obvious possibilities of leaking data, more sophisticated attack possibilities must be taken into consideration, also known as side channel attacks: For example, it is possible to intercept keyboards by the sound emanated when typing. By the execution of an acoustic

triangulation attack, whole sessions can be attacked with high recognition rates. Only publicly available tools and hardware is necessary, therefore, the attack can be performed even by non-technical people [35].

Because of this, adequate organisational measures must be conducted when COTS are used in applications relevant to security, e.g. the selection of inside rooms, measuring of the radiant emittance or consequent encryption of transmitted data.

In the context of high-security systems, software versions as well as system configurations must be released by the responsible competent authority. On the one side, these processes can be quite time-consuming, typically lasting several months or even longer. Therefore, the software products, e.g. operating systems are used for as long as possible during the life-span after the acceptance test and approval. On the other side, master agreements often do not include every new software release because of financial reasons, also introducing delays in the software regeneration. For this reason, the used software does not keep up with its life-cycle carried on by the manufacturer, resulting in out-dated and vulnerable installations in security-related systems.

One must also bear in mind that isolated systems and networks are no longer protected against attacks as examples like Stuxnet demonstrated. The weaknesses of human beings and today's sophisticated social engineering techniques [36] compromise even isolated and high-security systems. The successful attacks on RSA and the subsequent compromise of Lockheed Martin, Northrop Grumman and other companies of the American defence industry (e.g. see [37]) is only one example from recent years.

Several manufacturers of proprietary software have introduced so-called patchdays due to organisational and practical aspects, e.g. Microsoft, Oracle or Adobe (e.g. see [38]). On the other side, this policy unnecessarily delays patches, enabling crucial points of attack. Also, it is not guaranteed that the manufacturer will include all necessary patches, as the example of the thumbnail hole in Windows demonstrated: even though a Metasploit module for creating corresponding malicious files was released almost simultaneously with the security advisory of Microsoft, no patch was included in the subsequent patch day [39]. Another problem of proprietary software is the dependency to the vendor and his promises. For example, Microsoft announced to continue the support of Windows NT 4 until the end of 2004. Even so, the company stated they would not provide a patch for a new security vulnerability in NT 4 early in 2003 [40]. In contrast to FOSS, where it is always possible to fix an identified vulnerability, one is adhered to the vendor in the case of COTS.

Another aspect which must be mentioned in this context is what Bruce Schneier calls "bad civic hygiene". A rising trend in recent years is that governments force companies to redesign their communication systems and information networks to facilitate surveillance [41]. This is based on their desire to be able to pursue criminal activities. Even though this is a homemade problem, by introducing such backdoors, serious security vulnerabilities are opened up which can also easily be exploited by an attacker.

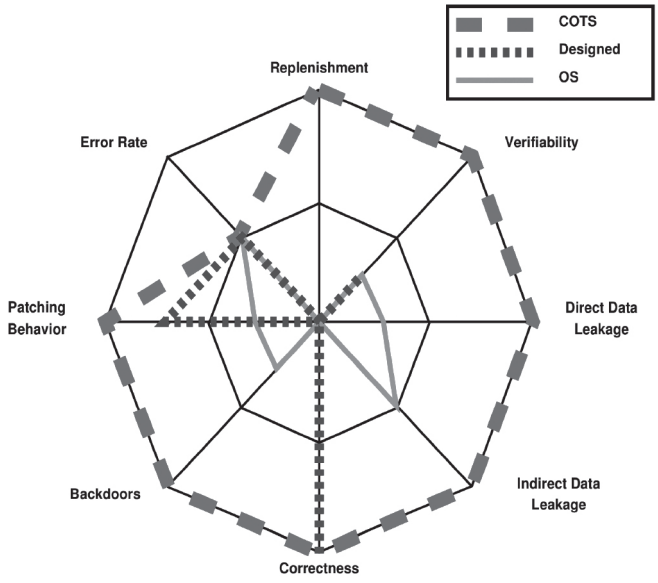
5. USING COTS PRODUCTS IN HIGH SECURITY SYSTEMS

Based on the identified influencing factors, the crucial aspects for using COTS, open-source and designed products are summarized qualitatively in Figure 1.

In detail, the following aspects must be taken into consideration:

- **Replenishment:** Especially for COTS products, the availability can be challenging after a few years. This cannot necessarily be compensated by storage because of the electronic components used. When using designed products, the supply can be governed by contract, typically reflected in high costs. Open-source enables the remanufacturing as needed; however, only a few circuits are available as open-source.
- **Verifiability:** While designed as well as open-source products can be verified with respect to their implementation, this is quite difficult for COTS.
- **Direct Data Leakage:** COTS products often implement undocumented functionality for statistical evaluation, etc. Also, a hardly detectable outward transfer of data can be integrated in COTS products.
- **Indirect Data Leakage:** Because of their cost-oriented design and fabrication as well as the fuzzy regulations, COTS products are strongly at risk of leaking data by radiation. Open-source can also be endangered by that phenomenon, but can be adapted and secured more easily. On the other side, designed products can be shielded per se.

FIGURE 1: INFLUENCING FACTORS ON SECURITY DEPENDING ON THE PARADIGM, *COMMERCIAL OFF-THE-SHELF, OPEN-SOURCE AND DESIGNED SYSTEMS.*



- **Correctness:** The public analysis and discussion of algorithms and procedures can reveal design errors in an early state. While Security-by-Obscurity can be quite effective in restricted military domains, it typically will not be in the public market and the use of widespread COTS.
- **Backdoors:** The difficult and limited test and control opportunities of COTS open up an endangerment by backdoors.
- **Patching Behaviour:** In the case of vulnerabilities, COTS depends on the manufacturer. Also when using designed products, later requests for patches can produce high costs. In contrast, fixing open-source can be quite easy due to the available code, even when there is no support.
- **Error Rate:** The error rates of all paradigms strongly depend on the design and development principles and techniques, and are not predictable.

To control the presented threats opened up by the use of COTS, several actions should be taken; on the other side, corresponding vulnerabilities in target equipment can be exploited for information operations in cyber space. The following aspects have to be considered:

- The communication of high-security systems should be statistically analysed to detect covert channels and unwanted behaviour. Because of the limited number of services in high-security networks, anomaly-based detection can be used to detect unwanted behaviour while achieving low false alarm rates. However, this may not be enough if a malicious behaviour is implemented from the beginning into a new device, because the correct traffic characteristic has to be known by the security system. Here, the use of unsupervised learning techniques can be an approach.
- Measurements of the radiation emittance must be done in areas where no adequate structural protection can be guaranteed by the buildings. It is important to include all possible media and connections, e.g. electromagnetism over the air, acoustics, interlinking in the power network, etc. While Tempest can be very powerful if cyber components are able to operate in the target area or adjacencies, the typical information operation will be conducted over long distances and, therefore, not able to exploit this valuable information.
- When using COTS in environments relevant to security, only long-term supported software and hardware should be used. Especially, only high-quality products should be purchased, including sufficient spare-parts. Suitable and controlled storage is a must-have for enabling adequate replenishment.
- If COTS are used in high-security systems, a doubling of systems can be used to strongly increase security while keeping costs reasonable. By the use and implementation of two independent products and the comparison of their calculations, anomalies and manipulations can be detected more easily.
- Where possible, COTS should be replaced by suitable open-source software and algorithms as well as open standards to be able to minimise design errors, etc.

Table 2 summarises important threats and attack opportunities related to COTS products.

TABLE 2: REQUIREMENTS FOR ENSURING SECURITY WHEN USING COTS PRODUCTS AND ATTACK POSSIBILITIES IN INFORMATION OPERATIONS RELATED TO COTS PRODUCTS IN THE TARGET ENVIRONMENT.

	Hardening	Information Ops.
Verifiability	High	Low
Direct Data Leakage	High	Medium
Indirect Data Leakage	High	Low/High
Correctness	Medium	Medium
Backdoors	High	High
Clearance	Medium	High

6. CONCLUSION

COTS products are used in ever more areas, for example for high-security systems and networks, and for hardware as well as software. By the use of COTS in areas relevant to security, numerous endangerments arise. Not only evident aspects like the lack of verifiability, but also secondary factors like replenishment and long-term availability must be taken into consideration. Therefore, the use of COTS products for mission-critical applications poses an imminent challenge. Even so, this endangerment is widely neglected at the moment: Based on the ongoing proliferation of attack tools and the numerous vulnerabilities opened up by the use of COTS, current and especially prospective military missions can be easily compromised: on the one side, effective attacks can be conducted even by an amateur. On the other side, aspects like reliability and supportability can strongly affect missions. With respect to the increasing financial pressure and the comprehensive use of COTS, it is crucial to address these challenges in depth. Therefore, an assessment of the usability and endangerment by the use of COTS in high-security environments must consider all layers in use, hardware as well as software. Based on the identified shortcomings, the high risk opened up by COTS can be attested. Appropriate countermeasures must be taken to overcome these endangerments, e.g. the statistical analyses of network communication. On the other side, an appropriate protection and examination of COTS can produce important knowledge about attack vectors, usable for own information operations in the cyber domain. Therefore, own system vulnerabilities must be identified and closed, and weaknesses must be known to keep superiority in information operations and to be able to defend from countermeasures.

ACKNOWLEDGEMENT

This work was done at the Chair for Communication Systems and Internet Services, led by Prof. Dr. Dreo Rodosek, part of the Munich Network Management (MNM) Team.

REFERENCES:

[1] NATO Research and Technology Organisation, "Commercial Off-the-Shelf Products in Defence Applications (The Ruthless Pursuit of COTS)", in *Information Systems and Technology Panel (IST-016)*, 2000.

[2] Federal Office for Information Security. (2007). *IT-Grundschutz Catalogues* [Online]. Available: https://www.bsi.bund.de/EN/Topics/ITGrundschutz/ITGrundschutzCatalogues

[3] T. Grance *et al., "Guide to Selecting Information Technology Security Products,"* NIST Special Publication 800-36, 2003.

[4] K. Scarfone *et al., "Guide to General Server Security,"* NIST Special Publication 800-123, 2008.

[5] *Common Criteria Common Methodology for Information Technology Security Evaluation,* Common Criteria Recognition Arrangement, 2009.

[6] *Security Target, Version 1.22 for XTS-400, Version 6.4.U4,* BAE Systems Information Technology, Inc., London, UK, 2008.

[7] G. Klein et al., "seL4: Formal Verification of an OS Kernel," in *ACM Symp. Principles of Operating Systems (SOSP)*, Big Sky, MT, 2009.

[8] M. Kuhn and R. Anderson, "Soft Tempest: Hidden Data Transmission Using Electromagnetic Emanations in Information Hiding," in *LNCS 1525*, 1998, pp. 124-142.

[9] M. Kuhn, *Compromising Emanations: Eavesdropping Risks of Computer Displays*, University of Cambridge, Computer Laboratory, Cambridge, UK, Tech. Rep. UCAM-CL-TR-577, Dec. 2003.

[10] M. Vuagnoux and S. Pasini, "Compromising electromagnetic emantations of wired and wireless keyboards,," in *Proc. 18th Conference on USENIX Security Symp. (SSYM09)*, USENIX Association, 2009, pp. 1-16.

[11] Z. Hongxin *et. al*, "Recognition of electro-magnetic leakage information from computer radiation with SVM," in *ScienceDirect Computers & Security*, no. 28, issues 1-2, pp. 72-76, 2009.

[12] Storelab. (2010). *(Comparison of the reliability of hard drives of the main producers)* [Online]. Available: http://www.storelab-rc.ru/sravnenie-nadezhnosti-hdd.htm

[13] *Directive 2002/95/EC of the European Parliament and of the Council of 27 January 2003 on the restriction of the use of certain hazardous substances in electrical and electronic equipment*, COD 2000/0159 extended to the EEA by 22003D0147, 2003.

[14] Czernobyl. (2010). *Super-secret debug capabilities of AMD processors* [Online]. Available: http://www.woodmann.com/collaborative/knowledge/index.php/Super-secret_debug_capabilities_of_AMD_processors_!

[15] G. Bloom *et al*., "Fab Forensics: Increasing Trust in IC Fabrication," in *IEEE Int. Conf. Technologies for Homeland Security (HST)*, Waltham, MA, 2010, pp. 99-105.

[16] G. Shah *et al*., "Keyboards and Covert Channels," in *15th USENIX Security Symp*., USENIX Association, 2006, pp. 59-75.

[17] V. Berk et al., *"Detection of Cover Channel Encoding in Network Packet Delays,"* Dartmouth College, Hanover, Tech. Rep. TR2005-536, 2005.

[18] S. Zander *et al*., "A Survey Of Covert Channels And Countermeasures In Computer Networkprotocols," *IEEE Commun. Surveys & Tutorials*, vol. 9, no. 3, pp. 44-57, 2007.

[19] K. Nohl and D. Evans, "Reverse-Engineering a Cryptographic RFID Tag," in *17th USENIX Security Symp*., USENIX Association, 2008, pp. 185-194.

[20] Dafyyd and Stuttard, "Security & Obscurity," in *ScienceDirect Network Security*, no. 7, pp. 10-12, 2005.

[21] J.H. Hoepman and B. Jacobs, "Increased security through open source", *Commun. ACM*, vol. 50, no. 1, pp. 79-83, 2007.

[22] The MITRE Corporation. (2012). *Common Vulnerabilities and Exposures* [Online]. Available: http://cve.mitre.org/

[23] cvedetails.com. (2012). *CVE Details: Windows 7 Vulnerability Statistics* [Online]. Available: http://www.cvedetails.com/product/17153/Microsoft-Windows-7.html?vendor_id=26

[24] P. Mell *et al*. (2007). *Common Vulnerability Scoring SystemVersion 2.0* [Online]. Available: http://www.first.org/cvss/cvss-guide.html

[25] R. Charpentier and M. Debbabi, "Security Evaluation and Hardening of Free and Open Source Software (FOSS)", in *Information Systems and Technology Panel* (IST-091), NATO Research and Technology Organisation, 2010, pp. 18-1–18-16.

[26] R. Panko. (2008). *Ray Panko's Human Error Website* [Online]. Available: http://panko.shidler.hawaii.edu/HumanErr/Index.htm

[27] J. Rutkowska, *Subverting Vista Kernel For Fun And Profit*, presented at the Black Hat Japan, Tokyo, 2006.
[28] S. Embleton et al., "SMM rootkits: a new breed of OS independent malware", in *Proc. 4th Int. Conf. Security and Privacy in Commun. Networks (SecureComm '08)*, 2008, pp. 11:1-11:12.
[29] D.K. Mulligan and A. Perzanowski, "The Magnificence of the Disaster: Reconstructing the Sony BMG Rootkit Incident", *Berkeley Technology Law Journal*, vol. 22, pp. 1157ff., 2007.
[30] US-CERT. (2010). *Energizer DUO USB battery charger software allows unauthorized remote system access (Vulnerability Note VU#154421)* [Online]. Available: http://www.kb.cert.org/vuls/id/154421
[31] NDJ World. (2012). *Secret Government Spying Build Into Smartphones* [Online]. Available: http://www.nodeju.com/17809/secret-government-spying-build-into-smartphones.html
[32] T. Eckhart. (2011). *Android Security Test: CarrierIQ* [Online]. Available: http://androidsecuritytest.com/features/logs-and-services/loggers/carrieriq/
[33] Directive 1999/5/EC, 1999.
[34] Directive 2004/108/EC, 2004.
[35] A. Hiu and Y. Fiona, *"Keyboard Acoustic Triangulation Attack"*, M.S. Thesis, Chinese Univ. of Hong Kong, Hong Kong, 2006.
[36] S. Gold, "Social engineering today: psychology, strategies and tricks", *ScienceDirect Network Security*, vol. 2010, issue 11, pp. 11-14, 2010.
[37] F.Y. Rashid. (2011). *Northrop Grumman, L-3 Communications Hacked via Cloned RSA SecurID Tokens* [Online]. Available: http://www.eweek.com/c/a/Security/Northrop-Grumman-L3-Communications-Hacked-via-Cloned-RSA-SecurID-Tokens-841662/
[38] R. Lemos. (2003). *Microsoft details new security plan* [Online]. Available: http://news.cnet.com/2100-1002-5088846.html
[39] H Security. (2011). *Microsoft warns of thumbnail hole in Windows* [Online]. Available: http://www.h-online.com/security/news/item/Microsoft-warns-of-thumbnail-hole-in-Windows-1163562.html
[40] P. Roberts. (2003). *Failure to Patch NT Flaw Causes Concern* [Online]. Available: http://www.pcworld.com/article/110054/failure_to_patch_nt_flaw_causes_concern.html
[41] B. Schneier. (2010). *Web snooping is a dangerous move* [Online]. Available: http://edition.cnn.com/2010/OPINION/09/29/schneier.web.surveillance

2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Conceptual Framework for Cyber Defense Information Sharing within Trust Relationships

Diego Fernández Vázquez,
Oscar Pastor Acosta
Defence and Security Division
ISDEFE
Madrid, Spain
{dfvazquez, opastor}@isdefe.es

Sarah Brown,
Emily Reid
Cyber Security Division
The MITRE Corporation
Bedford, MA 01730
{sbrown, ereid}@mitre.org

Christopher Spirito
International Operations
The MITRE Corporation
Bedford, MA 01730
cspirito@mitre.org

Abstract: Information and Communication Technologies are increasingly intertwined across the economies and societies of developed countries. Protecting these technologies from cyber-threats requires collaborative relationships for exchanging cyber defense data and an ability to establish trusted relationships. The fact that Communication and Information Systems (CIS) security[1] is an international issue increases the complexity of these relationships. Cyber defense collaboration presents specific challenges since most entities would like to share cyber-related data but lack a successful model to do so.

We will explore four aspects of cyber defense collaboration to identify approaches for improving cyber defense information sharing. First, incentives and barriers for information sharing, which includes the type of information that may be of interest to share and the motivations that cause social networks to be used or stagnate. Second, collaborative risk management and information value perception. This includes risk management approaches that have built-in mechanisms for sharing and receiving information, increasing transparency, and improving entity peering relationships. Third, we explore procedural models for improving data exchange, with a focus on inter-governmental collaborative challenges. Fourth, we explore automation of sharing mechanisms for commonly shared cyber defense data (e.g., vulnerabilities, threat actors, black/white lists).

In order to reach a common understanding of terminology in this paper, we leverage the NATO CIS Security Capability Breakdown [19], published in November 2011, which is designed to

1 The ability to adequately protect the confidentiality, integrity, and availability of Communication and Information Systems (CIS) and the information processed, stored or transmitted.

identify and describe (CIS) security and cyber defense terminology and definitions to facilitate NATO, national, and multi-national discussion, coordination, and capability development.

Keywords: *information sharing, cyber defense, framework*

1. INTRODUCTION

Information and Communication Technologies are increasingly intertwined across the economies and societies of developed countries. Protecting these technologies from cyber-threats[2] requires collaborative relationships for exchanging cyber defense[3] information and an ability to establish trusted relationships. The fact that cyber defense is an international issue increases the complexity of these relationships. Cyber defense collaboration presents specific challenges since most entities would like to share cyber defense data but lack a successful model to do so that takes into account the cultural perspectives of sharing and information exchange. We will explore the following four aspects of cyber defense collaboration to identify approaches for improving cyber defense information sharing:

- **Incentives and barriers for information sharing.**
  Aimed to identify the static structure of the information sharing network, and mainly trying to find answers of Why, Who and What of the network.
- **Information value perception and collaborative risk management.**
  Entities share information according to its perceived value, purpose, and meaning; thus, it is critical to ensure all entities have a common understanding of the information to be shared. It is critical to ensure all entities have a common understanding of the information to be shared. Depending on the nature and scope of the network, the approaches for collaborative risk management have to be shaped according to the prevention or response approach of the collaboration.
- **Improving data exchange.**
  Many cyber defense sharing networks suffer from an over-generalised concept of operations. Procedural models provide a structure that defines how information will flow across operational components. These models must address the information needs of the individual participants within each nation in order to provide sought-after information in a clear way. Bringing together information from complementary angles helps participants to derive results for problems that they cannot address individually.
- **Automation of sharing mechanisms for technical cyber defense data.**

A cyber defense information-sharing network is likely to contain a huge amount of technical data. Automation on the selection of that data and the mechanisms to share with participants

2 Threats are threat sources (or agents) with capability and intent, modeled as generic threats and specific threats. For example, Internet threats could be an instance of a generic threat and a certain hacker group could be an instance of a specific threat. Threat capability includes the ability of a threat source to perform certain activities such as using, customizing, and creating exploits, performing cryptanalysis, social engineering, etc. This can also include the various tools and resources that are available to the threat. This information can be tied to the CIS information for risk assessment. [12]

3 The ability to safeguard the delivery and management of services in an operational Communications and Information Systems (CIS) in response to potential and imminent as well as actual malicious actions that originate in cyberspace. [12]

in the framework of a specific network is a key requirement to facilitate effective analysis and sharing. Moreover, the existence of an automated exchange can provide an incentive for joining the trusted network; automation increases the benefit the parties involved by receiving data quickly and eases the process of contributing data to the network.

2. INCENTIVES AND BARRIERS FOR INFORMATION SHARING

There is a long history across the cyber defense community of establishing information sharing repositories, creating data-exchange standards, and finding the repositories underutilised.

There is a significant amount of research on approaches for information sharing. However, within the field of cyber defense, there is debate about:

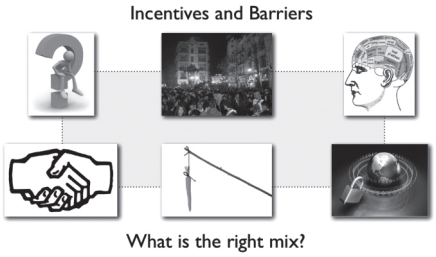
- Data types that are useful to share.
- Organizational and national policies about what can be shared.
- Models for sharing.
- How best to address privacy and security.

These questions, in which answers are still developing for the cyber defense community, add an additional challenge for sharing, because cyber defense is still not a well-defined, stable field. In addition to the maturity needed to determine what data to share and how to share it securely, more research is needed to understand social aspects of sharing. Engineers focus on technical aspects of information sharing networks, and often do not take into consideration the social, organizational, and cultural systems of use. In short, the motivations that cause communities to not engage in sharing or let a sharing relationship stagnate are not well understood. [10]

The European Network and Information Security Agency (ENISA) recently published a report on the barriers to and incentives for information sharing in the field of network and information security[1]. Taking these findings into account and to further our understanding of the motivations behind joining and participating in an information sharing community, we will explore the following:

- Why is the information-sharing network needed?
- Who will participate?
- What information is desired? What information will be shared/restricted?
- Does the network require services for confidentiality, integrity, privileged access and anonymity?
- What are the principles, challenges, and benefits in a cyber defense information-sharing network that will entice the right audience and achieve target objectives?
- Understanding incentives within information sharing networks
- Establishing, Perceiving and Maintaining Trust

FIGURE 1. I INCENTIVES AND BARRIERS INFOGRAPHIC



Why is the information-sharing network needed?
The network needs common scope and shared targets with the participants to reach the expected objectives of the information sharing from every participant. The scope specifies the approach – prevention, response or both - of the network.

Who will participate?
Once the scope and the objectives of the network are defined, the characterisation of the expected participant would be required based on organisational and individual aspects, for instance: the entity nature (public or private), network membership (mission or permanent), the scope of the organisation (national or supranational), and the functional role (technician or decision maker / governance). This information will allow for the creation of sharing profiles, used by sharing network participants to facilitate information exchange.

What information is desired? What information will be shared/restricted?
In addition to technical data, best practices and risk assessments may be of interest to share, attending to the role of the participants.

Does the network require services for confidentiality, integrity, privileged access, and anonymity?
The relationships between the participants need to be defined according to the requirements of the information to share. The specification of different scenarios will be necessary to consider the various options that may occur in the exchange of information to build trust between the players, either by the quality of information exchanged, authentication of its source, ensure the delivery of the information to authorised recipients or guarantee the anonymity of authorised participants.

What are the principles, challenges, and benefits in a cyber defense information-sharing network that will entice the right audience and achieve target objectives?
Entities participate in sharing networks when their return is more than the cost to participate. The identification of the benefits - for instance: cost savings, quality of information or network's relevance to the organisation - and the challenges - for instance: achievement of a high quality of information or establishment of clear and agreed management rules - of every potential participant will help to build the collaboration network and the principles that it is based on.

Understanding incentives within information sharing networks
The procedural model and its components must identify and use the incentives for sharing between participating entities. An assessment must be made of each participating entity type, their ability to produce products with perceived value, and the underlying incentives, such that the incentives can be threaded into the established sharing network procedures. Information economy aspects could be structured in financial incentive models that should be integrated into procedural models.

Establishing, Perceiving and Maintaining Trust
In an ENISA study of successful public private partnerships [6], one recommendation is about the importance of Trust Building Policies. The ENISA study reports that in information sharing networks where information sharing is the core service provided, a key requirement is a high degree of trust in the network itself (i.e., that the policies, membership rules, requirement for security clearance, and interaction type must have been carefully designed to support trust.

Trust between entities need not be whole or persistent. Transient trust during a moment of crisis may allow for a piece of information to be shared between two entities that would have not otherwise been made available for consumption. A sliding trust scale that is influenced by other factors such as operational need and quality of relationship must be incorporated into a sharing network to accommodate information sharing relationships that change in form over time. The partner you don't trust today may be your best friend tomorrow.

Trust relationships must span the different engagement levels: from the organisational leaders that empower their staff to produce and consume information to the technical staff that ultimately will take the information and put it to use. Having an institutional process for guiding these types of relationships is central to the success of an organisation as a whole in participating in information sharing networks. To support these processes organisations will need to focus on the trust scale while leveraging mechanisms and tools to support the mapping and perception of these relationships.

Trust relationships are affected by both the organizational and ethnic cultures of the sharing entities. There are cultures where no information sharing will take place until a maturity point is reached in the relationship. Then there are ethnic cultures where a business need will drive information sharing even though the relationship has not matured enough for sustained information sharing between entities.

3. INFORMATION VALUE PERCEPTION AND COLLABORATIVE RISK MANAGEMENT

Entities share information according to its perceived value, purpose, and meaning; thus, it is critical to ensure all entities have a common understanding of the information to be shared. At the human and machine level, establishing trust and effective communication requires a common vocabulary and taxonomy, especially between nations with different languages. For example, in this paper, we refer to the NATO CIS Security Capability Breakdown [12] to ensure

a common understanding of CIS and cyber defense terminology that appears. The CIS security capability breakdown is designed to specifically facilitate NATO, national, and multi-national discussion, coordination, and capability development related to CIS security and cyber defense.

When we look further into how entities view of particular piece of data or situation, we find this topic explored by "ethnomethodologists", who use the phrase "sense-making" to refer to observable behaviours in which individuals orient toward the same aspect of the world and demonstrate to each other – through detailed enactment of practices – that they share that orientation. "Mutual orientation toward an object" includes:

- Perception (we're looking at the same thing),
- Interpretation or instructed perception (we're looking at the same aspects of, or applying the same framework on, that thing), and
- Conventions or instructed actions (we display similar behaviours with respect to use of that thing; the modifier "instructed" refers to the fact that we learn those behaviours from on another, primarily by example)." [2]

This first step in the analysis of an information sharing relationship is critical, especially when two or more countries and cultures are involved. There must be an agreement from all parties that the shared perception of the objects in the repository exists. The second step is to ensure that all parties agree upon the analysed characteristics of the framework. Lastly, there needs to be an ability to include the behavioural components of information sharing so that acceptable boundaries are placed around. Standards ensure entities agree on the information to share and can exchange it.

Assessing and mitigating existing risks is easier than anticipating unknown risks. Thus, risk management approaches should include collaborative models with built-in mechanisms for sharing and receiving information, increasing transparency, and improving entity peering relationships. These approaches should facilitate government relationships and public-private partnerships.

Traditional risk management usually consist of two phases, no matter what is the applied methodology such as NIST SP800-30 [3], ISO 27005 [4], or MAGERIT [5], aimed to gather the risk awareness in a specific time that has to be updated– usually yearly - in a regular basis:

- risk assessment that could be generally described as an identification of assets, threats and countermeasures to obtain assessments of the risk stemming from the impact on the assets
- risk management where it takes into account the risk assessment to make decisions on how every identified risk will be managed.

In case that the information sharing network is focused on the prevention approach, the information flow should be related to preparation against threats that can exploit vulnerabilities causing impacts on assets. Sharing of new or evolved vulnerabilities, patterns of threats, new or evolved threats, technical countermeasures and non-technical countermeasures are expected.

In case that the information sharing network is focused on the response approach, the information flow should be related to how the risk is managed mainly in the response to and recovery from the attacks based on the impact. Sharing of how the collaboration could be more efficient, how mutual aid agreements could be adopted, identification of cascading effects, practices to improve the efficiency on the recovery of services, operational responses to attacks and collaboration procedures are expected.

But there will be a subjective factor on the risk management because of the diverse rules or perception on definitions of threat levels, identification of relevant assets, identification of countermeasures to apply and how the impact is considered as relevant in organisations. Organisations could come from diverse cultures/sectors (the principal assets to protect) and countries (diverse languages could cause difficulties since translated words and sentences may not have the exact or equivalent meaning) that could produce some misunderstandings on how the risk is managed within an environment of aggregated risk management where cascading effects have to be avoided and the trust among participants of the sharing network needs to be held or improved to foster their collaboration.

As the situational awareness of the cyberspace related to an organisation is in a very changing environment, a specific organisation can take data related to the status of cyber defense in order to calculate in real time the threat level and share with participants of its collaboration network. An agreement on how the threat level is calculated and the meaning of each threat level – in terms of expected impact and expected actions of reaction - is envisaged as a mandatory pre-requirement for collaborations based on mutual understanding of the different risk management approaches. This could support a dynamic risk management where threat levels are calculated in real time, as opposite to traditional risk management, and providing the appropriated information to decision makers about how the risk have to be deal with – updating the threat awareness support a quick, efficient and adaptable reaction to the changing attack environment - and how to anticipate risk to selected participants – for instance based on mutual aid collaboration agreements - of the collaboration network.

4. PROCEDURAL MODEL FOR IMPROVING DATA EXCHANGE

Many cyber defense sharing networks suffer from an over-generalised concept of operations. Procedural models[4] must dictate how information will flow across operational components so that flows can be optimised and information products can be integrated into decision trees.

Information exchange models must address the information needs of the individual participants within each nation in order to provide sought-after information in a clear way. The data sharing network should bring together information from complementary angles, allowing participants to derive results for problems that are difficult to address individually. Aspects that must be considered to design effective procedural models for a cyber defense sharing network include:

4 The generally simplified representation of an aspect of reality expressed in a specified manner so as to facilitate reasoning about that aspect.[12]

- Participant Roles
- Governance Structure
- Institutional Funding
- Enabling Collaboration
- Information Protection and Release Control
- Incorporating Financial Incentive Models

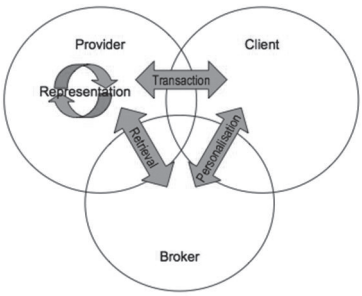
Participant Roles

We know from experience that the value of information varies based upon the needs of the consumer. Each information consumer assigns values to the types of information they need in the moment. Each information producer assigns a value or cost for the piece of information they are sharing. A successful information sharing network will bring together information producers and consumers with minimal friction. To achieve this, each participant must be assigned a role for a specific transaction. Participants may act in various roles within the information sharing network, but for any transactions, we must be able to define the role held by each participant in that transaction.

When we talk about participants, we are not limited to participants as individuals. Rather we are taking the view that a participant can be a non-organisationally associated individual on one end of the spectrum, or a multinational entity that has multiple types of participants within it at the other end of the spectrum. We do exclude non-human participants such as Artificial Intelligence backed systems.

Roland Klemke in his Modeling Context in Information Brokering Processes thesis states that "three different roles participate in the information brokering process: the provider who offers information, the consumer who demands information, and the broker who mediates between the other two. Different roles in this view not necessarily have to be represented by different persons, a role may even be represented by fully automated processes." [16] We also include the role of Information Producer as we recognise in the world of cyber security the producer of information may often not be the provider offering the information to a community.

FIGURE 2. THE SEMANTIC WEB WITH INFORMATION BROKER.



Participant roles within a transaction include:

- Information Producer - the entity that has drafted a piece of information for publication
- Information Provider - the entity that is publishing the information to the repository. This may not always be the same as information producer in the cases where the producer would like to stay anonymous
- Information Consumer - all entities that have consumed a piece of information.
- Information Broker - an entity that negotiates between two or more entities arranging for the publishing and consuming of information

> "Information brokering is a pragmatic means of knowledge exchange: ..., knowledge cannot be exchanged directly. However, knowledge can be externalised and re-conceptualised (i.e. transformed into information) and then exchanged as information. At the receiving party, the delivered information can then be turned into knowledge by contextualisation again." [16]

Clearly defining participant roles allows for a bounded exchange of information, holding each participant to pre-defined rules when acting in that role within the defined cyber defense sharing network.

When describing an Information Broker, an organisation may explicitly choose to be a primary information broker within a network so that it gains the widest and deepest view of network knowledge. However, organisations may only become a trusted information broker when the level of perceived trust with that organisation is sufficiently high enough across participating organisations such that that organisation brokers the flow of information between participants that do not have a high enough perceived trust between each other.

Governance Structure

The governance structure of a cyber defense sharing network within an information sharing environment must address two distinct areas:

First, there is the governance structure of the network participants:

- how participants are structured (e.g., defined roles and responsibilities)
- what are the duration, participation and interaction types,
- what sharing network membership and usage rules are in place to handle day to day activities and address dispute resolution between participants,
- what kind of trust-building policies are in place to encourage success.

Governance also addresses the information sharing relationships between participating entities. Specifically, it is a description of the top cover needed by sharing entities to ensure each producer and consumer is empowered by their management to share specific types of information.

FIGURE 3. PPP INFOGRAPHIC FROM ENISA PAPER

Using ENISA's publication on Preliminary Taxonomy for Public-Private Partnerships (PPPs) [6] as a guide to governance structure, we will walk through each component, specifically focusing on the incentive impacts for each component.

Organisation

The ENISA paper [6] references the Milward and Provan model on collaborative networks where all networks are describable using three constructs: run by one from within, run by a coordinating entity, and democratically peer led. We have conducted an initial set of interviews with members of two incident response teams and our preliminary research indicates that the most successful cyber defense information sharing model is the democratically peer led network where individual trust relationships tend to increase the amount of sharing that takes place. From what we have also observed partnerships that have a "run by one from within" structure tend to form more quickly but later fail to gain traction.

Roles and Responsibilities

The roles and responsibilities within an information sharing network can be non-exclusively tagged to these taxonomy categories: Chaired by {elected representatives from Industry, representative from Government}, Secretariat supplied by {third party (non-government), national government}, and Co-ordinated by {government, industry and collectively). When the information-sharing network is very large, roles and responsibilities help to organise the community and maintain a common understanding of relationships and expected contributions from participants. Roles and responsibilities also help to clarify the goals of each participant for the community.

Duration Type

Governance structure and institutional funding are both impacted by the duration type of the sharing network. Some sharing networks are classified as persistent community groups, setup to serve a community of interest without a bounded endpoint. A second classification bounding the duration type of a sharing network is a working group where specific problems are addressed and the group is disbanded once objectives are met or the group is disbanded. The third duration type classification is a rapid response group that is more or less an extension of the working

group in that the sharing network is created to address an urgent issue and may only be in existence for a matter of hours or days.

Participation Type
Participation dynamics within sharing networks are interesting from the perspective of both corporate governance as well as individual motivations. A successful sharing network may only succeed by providing entry-points for all types of participants. Participation can be in the form of a subscription where a participant pays a fee (or just subscribes) to a sharing network to gain access to the collective knowledge. While subscription based services describe a mechanism for interacting with sharing networks, two other participation types describe a commitment level for participants, either mandatory or volunteer. Mandatory participation may be leveraged upon an individual or organisation by the owning entity such as a government. Voluntary participation may, on the other hand, incentivise a participant to use the information sharing network since they may wish to shape their participation based upon their organisational or operational priorities.

Interaction Type
The ENISA PPP paper [6] outlines two interaction types: face-to-face and virtual cooperation. This is largely an extension of the time/place collaboration square where sharing mechanisms vary according to the location of participants and the length of interaction. Governance structure will often dictate the interaction type but successful interaction within a cyber information sharing network will often be based upon the duration type (severity of engagement).

Formal Information Usage Agreements
Information which is shared in a cyber defense sharing network must be protected. This requires a legal component – who is the information owner, how can the information be used, can it be attributed to the owner, etc.

Trust Building Policies
Building trust has two components. First, participants will develop trust in the cyber defense sharing network as participants feel that the information they contribute is protected (e.g., the network should be able to provide anonymisation for contributed data), and that the network provides them the opportunity to gather valuable information unavailable elsewhere, providing high value back to participants (e.g., bringing in participants with expertise that incentivise new members).

Second, participants will develop trust in each other over time as their relationships strengthen. In our experience, holding face-to-face meetings throughout the year significantly increases trust building among participants. Highlighting shared goals and facilitating partnerships among participants to realize these goals will also go a long way to building strong trust and partnership in a cyber defense sharing community.

Establishing Collaborative Processes
The multi-dimensional view of information sharing transactions requires a defined collaborative process. This defined process also helps to alleviate the anxiety of a transaction by providing to

each party a set of steps, responsibilities and time-to-act deadlines to facilitate the information exchange.

Information Protection and Release Control

Often we will see information sharing partnerships fail not because the two parties do not trust each other to have the information, but one party may doubt the other party's ability to protect information consumed appropriately. This is especially true in the case of classified data that may pass between nations or cyber threat signatures that if an adversary knew existed would allow for crafting of attack payloads that do not trigger (at least for that rule set) an alert.

The procedural model must include steps for protecting information as it is created, published, consumed, stored and eventually destroyed. The information exchange platform must itself be capable of protecting all information it stores from unauthorised access.

FIGURE 4. IPRC INFOGRAPHIC

Incorporating Financial Incentive Models

Within the malicious software community exploits are bought and sold based upon the perceived value of the exploit. Is it something that no one else even knows about? Does it affect a piece of software used by your targets? Is the author someone you can trust to have not sold the exploit to anyone else already? Does the asking price match the perceived value? Existing research, for example [18], shows that information-sharing networks need to incorporate these types of financial incentive models into their procedural underpinnings. Approaches as Worldwide Intelligence Network Environment (WINE) [7] could help to build financial incentive models. Not every network participant will bring the same capabilities to the table, therefore there may need to be a financial incentive in place in lieu of reciprocal information exchange such that those who have valuable information to share aren't vested because their return is not sufficient.

5. AUTOMATION OF SHARING MECHANISMS FOR TECHNICAL CYBER DEFENSE DATA

The need for automation and standardization of cyber defense data is apparent in the government, academic, and industry sectors on an international level. Information sharing that can relieve the human workload is necessitated by the sheer speed of cyber threats today. Standardization of data to be exchanged provides an effective pathway for information sharing between multiple parties, because the format of the data is then agreed upon. Standardization

also lends itself to automation of information sharing, and both lower the bar for entering into a cyber defense data sharing network.

Trust is a very important component in regards to automated information sharing. When the speed at which data could be shared increases, the risk of sharing information with unauthorized parties is raised, potentially backfiring and creating a disincentive for participation in an information sharing network. Nonetheless, the existence of an automated exchange can provide an incentive for joining the network; automation increases the benefit the parties involved by receiving data quickly and eases the process of contributing data to the network.

Additionally, the details of the sharing relationships and the automation involved depend heavily on the type and sensitivity of the information to be shared. Some information types are considered high-risk in sharing environments; they would reveal too much sensitive data and existing initiatives are faced this challenge as Sharemind [7,9]. Low-risk data, or data of less sensitivity, is more likely to be shared in an automated information exchange. It is important to keep in mind that the level of trust of the partners and the level of sensitivity of the data are directly related.

The data in a cyber-information sharing network could include the following types:

- Vulnerability information
  - Vulnerability existence checks
  - Related patches and mitigations
  - Quality of service effects
  - Vulnerability Assessment tests/results
- Threat actors
  - Names/pseudonyms
  - Countries of origin
  - Common methods and tactics
  - Attack patterns
  - Events and incidents
  - IDS Signatures
  - Implicated parties
- Black or white list information (IP addresses)
- Software
- Hardware
- Malware
- Protocol specifications
- Security configurations
- Security guidance
- Weakness information, patch remediation
- Secure coding practices

Of the above types, high-risk data may include specific threat actor information, especially attack patterns and methods. Internal security configurations are also high-risk. This is because

they can reveal sensitive information about the organization and may be shared with a party that is not trusted with that level of sensitivity. However, blacklist information, security guidance, or patch information may be considered lower risk, and are appropriate for an automated exchange without an exceptionally high degree of trust. Information sharing networks and the number of participants actively involved will most likely be directly related to the amount of data available. Since high-risk information is less likely to be shared, a low-risk sharing environment may create the best incentive for participation.

One example of an automated cyber defense-sharing network (including exchange of many data types) is CDXI (Cyber Defense Data Exchange and Collaboration Infrastructure) for Cyber Defense data exchange, a system being built by NATO [14]. CDXI will serve as a repository for participants worldwide (individuals, organizations, non-NATO entities, industry, government, and academia) that will automatically push and pull cyber defense data using a variety of Application Programming Interfaces (APIs). Quality assurance of data and data confidentiality are integral to the CDXI design, and in order to achieve the right balance of information protection (i.e., sharing with appropriate parties) and openness of the network, confidentiality and access control are implemented based on user, role, and NATO classification level.

CDXI data is to be structured for machine processing and automation but also have a human-readable component. Automatic exchanges exist for some of these information types, however in practice much of this information (such as configuration information and operational events) is exchanged via prose documents and requires manual interpretation and implementation. Automating the exchange of this data should likely increase efficiency, which not only increases the incentive to share and participate in the information sharing network, but also saves valuable time in securing an organization against fast-acting threats.

Automation, however, requires standardization of data before it can be automatically exchanged. An agreement between parties on the format of data is often required in order to exchange, so standardization in and of itself can provide an incentive for information sharing. One popular example of a data standardization protocol is the Security Content Automation Protocol (SCAP). SCAP includes a suite of standards that provide a common way to identify vulnerabilities (Common Vulnerabilities and Exposures or CVE), platforms (Common Platform Enumeration or CPE), and configurations (Common Configuration Enumeration or CCE), as well as a common way to express configuration information and security guidance (eXtensible Configuration Checklist Description Format or XCCDF), system configuration and vulnerability assessment (Open Vulnerability and Assessment Language or OVAL), and vulnerability risk (Common Vulnerability Scoring System or CVSS). These internationally accepted security standards encapsulate valuable vulnerability information and are widely used across government, academia, and industry.

The National Vulnerability Database or NVD is a freely accessible repository for SCAP data such as NVD contains CVE vulnerability feeds with CVSS scores, the CPE product dictionary, CCE reference data (and soon a vulnerability feed), and NCP (National Checklist Program) checklist feed. These checklists are usually a bundle of data including at least an XCCDF-expressed checklist, but also may be annotated with CVEs, CPEs, or CCEs and may include

OVAL definitions or other automated checking mechanisms. Each of these feeds is available in an RSS or XML format.

The NCP checklists are presented in tiers. The most important tiers for automated standardized data are Tiers 3 and 4. Tier 3 designates data that should work in an SCAP-validated tool (i.e., passes SCAP data stream requirements but may need to be tested), and Tier 4 designated data that does work in an SCAP-validated tool (i.e. passes SCAP data stream requirements and has been tested). While the contributors to these tiers have been primarily been government or government-contractor organizations (e.g. NSA, DISA, MITRE) there are a few examples of private companies that have adopted SCAP data formats and contributed content, forming a public-private partnership. Microsoft has been very involved in expressing its configuration information in the SCAP format. For example, Microsoft's SCM (Security Compliance Manager) now provides extensions to express configuration information in SCAP format. Additionally, Microsoft provided the Tier III Checklists to the NVD on a total of 12 platforms, including several versions of Windows operating systems, Office, and Internet Explorer. CyberESI is another private company that has contributed to the National Vulnerability Database using SCAP-formatted data. CyberESI is an information security company that provides services to both government and commercial clients. CyberESI developed a Tier 3 checklist that checks for suspicious filenames and locations on a Windows XP system. While they have not contributed to the NCP, Red Hat now includes in all of their security updates with OVAL definitions that check for the vulnerability or configuration issue. These are only a few of the major private contributors that have shared information in the standardized SCAP format.

In terms of information sharing networks, these databases provide an automatic yet mostly one-way trusted flow of information. While it is two-way in the sense that community members (which include government, academia, and industry) may provide the information to be vetted by NIST or MITRE, it is one-way to the largest population of users: the public. Since these websites are public and the total community of users is not controlled, they lack some of the ideal characteristics for a highly utilized information sharing network. However, the automatic ability to pull data in each case account for both repositories' reputation in the field of vulnerability and security configuration data, and may indirectly contribute to the volume of data (49,000+ CVE IDs, 7500+ OVAL queries and 220+ checklists) by creating a strong community of users. The important lesson to learn from these repositories is that when many parties, with many different ways of describing and expressing their data are trying to exchange non-standard information, the information can't be normalized. An important issue to consider, however, is how standardization is applied. For example, the success of CVE spawned the growth for many more security-related standards, but few have the widespread success that CVE did. Research [5] that examined why some standards are more successful than others found that differences between machine- and human-oriented standards contributed to a standard's success, and that this must be considered when using or developing standards for information sharing environments. In particular, standards that include little detail (e.g. a CVE ID), allow for a greater degree of diversity in the information represented, while a very detailed (i.e. more constraining) standard will result in very similar enumerations. This is an important consideration depending on the type of data to be shared in a particular environment.

Repositories with more sensitive information require collaborative trust to incentivize potential new users. One example is the U.S. Defense Security Information Exchange (DSIE). DSIE is an information exchange network for U.S. Defense Industrial Base (DIB) companies to share information on cyber-related events and attacks, formed in 2008 [12]. In order to facilitate sharing, DSIE members sign a Non-Disclosure Agreement (NDA) which states that all information is non-attributional and that only DSIE members can view the information.

Cyber information sharing networks with high participation will ideally contain a large amount of data. The collection, processing, and distribution of this data in the network are time consuming if done primarily manually. Automation of the exchange data is important to consider in the network. Automation may increase the incentive to join the network, share information, and continue to be an active user. Standardization plays an important role, since it is a prerequisite to data automation in some way. How standardization is used and applied depends on the data to be shared and its usage. Other considerations include the risk-level of automatically shared data and pre-existing trust relationships. While the technology and procedures around standardized and automated cyber information sharing must be carefully considered, standardization and automation ultimately provide a great incentive for sharing by reducing manual work and increasing efficiency.

6. CONCLUSIONS

Research into the field of incentive networks, specifically collaborative scenarios for sharing information within trust relationships, is still quite new. Throughout this paper we have presented a common sense approach for thinking about how incentives in sharing networks work. We started with identifying incentives and barriers for information sharing. We looked at the importance of modelling the networks for information sharing (the aim of the network, the goals of the participants, and the envisaged benefits and challenges of each participant to establish the principles and the procedures that rules the network) and then moved onto the idea of collaborative risk management models and the important notion of information value perception.

Once a clear common understanding is achieved with regards to these kind of networks, procedural models for improving data exchange will help to start driving an organisation towards integrating their risk models with their information sharing models such that an agreement of threat level, envisaged impact, risk methodology and finally mutual aid from a risk management point of view will help to improve the effectiveness of the collaboration network.

Over the next few months we will continue our research into sharing networks and incentives with the intent on providing a more thorough review of our research at the CyCon 2012 Conference this June in Tallinn.

REFERENCES

[1] ENISA. “Incentives and Challenges for Information Sharing in the Context of Network and Information Security”. September 2010.

[2] Bodeau, D., Powers, E., Brooks, J. Making Sense Together: Applying Ethnomethodology to Enhance Advanced Systems Engineering in the Information sharing Domain.

[3] NIST sp-800-53 Recommended Security Controls for Federal Information Systems. April 2009. Retrieved January 2012 from http://csrc.nist.gov/publications/nistpubs/800-53-Rev3/sp800-53-rev3-final_updated-errata_05-01-2010.pdf

[4] ISO/IEC 27005 Information technology - Security techniques - Information security risk management

[5] MAGERIT: Metodología de Análisis y Gestión de Riesgos de los Sistemas de Información. Retrieved January 2012 from http://administracionelectronica.gob.es/?_nfpb=true&pageLabel=PAE_PG_CTT_General&langPae=es&iniciativa=184

[6] ENISA. “Cooperative Models for Effective Public Private Partnerships. Desktop Research Report”. Retrieved January 2012 from http://www.enisa.europa.eu/act/res/other-areas/national-public-private-partnerships-ppps/desktop-reserach-on-public-private-partnerships/at_download/fullReport

[7] Dumitras et al. in “Toward a standard benchmark for computer security research: the worldwide intelligence network environment (WINE)”, BADGERS’11, 10 April 2011, Salzburg, Austria.

[8] Bogdanov, Dan., Laur, Sven., Willemson, Jan. Sharemind: a framework for fast privacy-preserving computations. In Proceedings of 13th European Symposium on Research in Computer Security, ESORICS 2008, LNCS, vol. 5283, pp. 192-206. Springer, Heidelberg (2008).

[9] Talviste, Riivo. Deploying secure multiparty computation for joint data analysis — a case study. Master’s thesis. University of Tartu, 2011.

[10] Mann, D., Brookes, J. Information Standards and Their Use: Implications and Design Patterns. March 2010.

[11] P. Welsh, “Newest version of DCGS Integration Backbone improves intelligence sharing”. Retrieved January 2012 from http://www.afmc.af.mil/news/story.asp?id=123228659

[12] “Defense Security Information Exchange (DSIE) A partnership for the Defense Industrial Base”. Retrieved January 2012 from http://www.whitehouse.gov/files/documents/cyber/Defense%20Security%20Information%20Exchange%20-%20DSIE%20summary%20-%20William%20Ennis.pdf

[13] ENISA. “United Kingdom Country Report”. May 2011. Retrieved January 2012 from http://www.enisa.europa.eu/act/sr/files/country-reports/UK.pdf

[14] L. Dandurand, “Cyber Defense Data Exchange and Collaboration Infrastructure (CDXI)”. ITU-T Workshop December 2010. Retrieved January 2012 from www.itu.int/dms_pub/itu-t/oth/06/35/T063500000200516PPTE.ppt

[15] NATO Consultation, Command and Control Agency Reference Document RD-3060, “CIS Security (Including Cyber Defense) Capability Breakdown”, G. Hallingstad, L. Dandurand, NC3A, The Hague, Netherlands, November 2011 (NATO Unclassified).

[16] R. Klemke. “Modelling Context in Information Brokering Processes”. Retrieved January 2012 from http://darwin.bth.rwth-aachen.de/opus3/volltexte/2002/381/pdf/Klemke_Roland.pdf

[17] The semantic web with information broker. Retrieved January 2012 from http://www.semanticweb.org/

[18] Golle, P., Leyton-Brown, K., Mironov, I., and Lillibridge, M. “Incentives for Sharing in Peer-to-Peer Networks”. Proceedings of the 3rd ACM Conference on Electronic Commerce, New York, NY. 2001.

[19] G. Hallingstad, L. Dandurand, NATO Consultation, Command and Control Agency Reference Document RD-3060, “CIS Security (Including Cyber Defense) Capability Breakdown”, NC3A, The Hague, Netherlands, November 2011 (NATO Unclassified).

BIOGRAPHIES

Editors

Christian Czosseck is a German Army Officer with more than 14 years of experience in information assurance and IT management, currently assigned as a scientist to the NATO CCD COE in Tallinn, Estonia. Christian, a distinguished graduate in Computer Science from the military university in Munich, is currently a PhD student at the Estonian Business School in Tallinn. His current research focus is on cyber security and cyber weapons, with a particular focus on the role of botnets. He has edited numerous conference proceedings and frequently serves as a programme committee member for cyber-related conferences.

Dr **Rain Ottis** is a scientist at the NATO Cooperative Cyber Defence Centre of Excellence in Tallinn, Estonia. He previously served as a communications officer in the Estonian Defence Forces, focusing primarily on cyber defence training and awareness. He is a graduate of the United States Military Academy (BSc, Computer Science) and Tallinn University of Technology (PhD, MSc, Informatics). His research interests include cyber conflict, politically-motivated cyber attacks and the role of volunteers in cyber security.

Dr **Katharina Ziolkowski** is a member of the legal service of the German Ministry of Defence, currently serving at the NATO CCD COE. Since she joined the legal service, Dr Ziolkowski served as the legal advisor to the project 'InfoOp and CNO Capabilities', as legal advisor and military prosecutor for the Airborne Operations Division of the German Armed Forces and as legal advisor and law lecturer at the NATO School, the US Army's JAG School and the German General Command and Staff College. Dr Ziolkowski studied law and political science in Berlin and Barcelona and is admitted to the German bar. She holds a law degree from Freie Universität Berlin, an LL.M. in International Law from the University of New South Wales (Sydney) and a *magna cum laude* doctorate from Freie Universität Berlin, awarded by the Friedrich-Naumann-Foundation and the German Society for Military Law and the Law of War.

*Authors**

Major **Scott D. Applegate** is a US Army Information Systems Management Officer with more than 20 years of leadership, management, communications and security experience. Major Applegate holds two Master's Degrees, one in Military Studies and one in Information Technology and Assurance, and is currently pursuing a PhD in Information Technology with a focus on Cyber Conflict and Information Security at George Mason University in Fairfax, Virginia. His current research interests include cyber conflict, cyber militias, security metrics, cyber security policy, information assurance and cyber law. Major Applegate currently resides in Northern Virginia with his wife Sara and their two children.

Dr. **Louise Arimatsu** has been an Associate Fellow with the International Law programme at Chatham House since 2009. She was also a visiting lecturer at UCL from 2009-2011, where she taught Foreign Relations Law. Between 2006 and 2009 she was a part-time lecturer at the

* Authors who have presented a biography

LSE where she taught International Criminal Law, International Law of Armed Conflict and the Use of Force, Refugee Law and Human Rights Law. She was also the Co-ordinator of the International Humanitarian Law Project at the LSE. Louise was Managing Editor of the Yearbook on International Humanitarian Law from 2009-12.

Sarah Brown has been Lead Cyber Security Engineer for the MITRE Corporation since 2004. For the past four years she and her family have lived in The Netherlands where she serves as a US technical liaison to the NATO C3 Agency in The Hague. Her current work focuses on cyber defence information sharing, coalition operations, cyber exercises and technical security standards. She particularly enjoys working across organisations, in a multinational environment, addressing common technical and strategic challenges. Prior to MITRE, Sarah held internship and fellowship positions at the National Academies, Sandia National Laboratories and the US National Security Agency. Sarah received bachelor degrees in Mathematics and Computer Science from Oberlin College and a Master's degree in Mathematics from University of Maryland, College Park.

Jeffrey Caton is Associate Professor of Cyberspace Operations at the US Army War College, supporting cyberspace curriculum development as well as conducting research, publications, doctrine reviews, and lectures. He also provides subject matter expertise in space, missile and nuclear operations. Prior to his current position he was an instructor in the college's Department of Command, Leadership and Management for three years, where he also served as the college's Defense Transformation Chair. Professor Caton served 28 years in the US Air Force with duties including space and missile operations, joint special operations and planning, and engineering, test, and programme management.

Myriam Dunn Cavelty is a lecturer for security studies, a senior researcher, and head of the Risk & Resilience Research Group at the Center for Security Studies (CSS), ETH Zurich (Swiss Federal Institute of Technology). She studied International Relations, History, and International Law at the University of Zurich. She was a visiting fellow at the Watson Institute for International Studies (Brown University) in 2007 and fellow at the Stiftung Neue Verantwortung in Berlin, Germany from 2010–2011. Her research focuses on the politics of risk and uncertainty in security politics and changing conceptions of (inter)national security due to cyber issues (cyber-security, cyber-war and critical infrastructure protection). In addition to her teaching, research and publishing activities, she advises governments, international institutions and companies in the areas of cyber security, cyber warfare, critical infrastructure protection, risk analysis and strategic foresight.

Colonel **David T. Fahrenkrug** is the Senior Military Assistant in the Office of Net Assessment. He conducts comparative analysis and comprehensive assessments on future military competition and conflict in cyberspace. A command pilot, he has flown combat missions in multiple operations. He holds a Doctorate in Political Science from the University of Chicago and was the lead author for the Air Force's concept document on warfare in cyberspace. Most recently, he was the Director of the Air Force's Strategic Studies Group, providing strategic assessment and recommendations to the Chief of Staff concerning military strategy and USAF contributions to national security.

Vittorio Fanchiotti is Professor of Criminal Procedure at the Law School of the University of Genoa; since 1985 he has also taught Comparative and International Criminal Procedure. In 1998, as a member of the Italian Delegation, he participated in the UN Diplomatic Conference on the Establishment of an International Criminal Court. From 1998-2000 he was a member of the UN Preparatory Commission, responsible for drafting the Rules of Evidence and Procedure of the abovementioned Court. He has participated in European projects on distance-learning for former Yugoslavian and Albanian judges and prosecutors on the topic of the European Convention for Human Rights. In 2005 he was member, on behalf of the European Commission, of the Identification Mission in Beijing (PRC) for the creation of the first Sino-European Law School (which was subsequently instituted). His fields of study, in which he has published both books and articles, are comparative criminal procedure, international criminal justice and human rights.

Robert Fanelli is a United States Army officer currently serving with the United States Cyber Command. He holds a BSc from the Pennsylvania State University, a MSc from the University of Louisville and a PhD in Computer Science from the University of Hawaii. Prior to his assignment at Cyber Command, he served as an Assistant Professor of Computer Science at the United States Military Academy at West Point. His duties included directing the Academy's Cyber Security course and serving as head coach of the West Point team in the annual inter-academy National Security Agency Cyber Defense Exercise.

Mr. **Diego Fernandez-Vazquez** has a degree in Telecommunication Engineering from Universidad de Vigo, and is CISA-certified, PRINCE2 Foundation-certified and ITIL Foundation-certified. He has been employed at ISDEFE since 2002, working as a security consultant on projects related to information assurance. He has worked mainly on European R&D projects and projects for the Ministry of Defence. As an adjunct teacher of cryptography at the Antonio de Nebrija University on academic courses from 2002-2003 and 2003-2004, he has participated in the European Security Research and Innovation Form (ESRIF).

Keir Giles: Following a first degree in Russian and one of the briefest careers on record in the Royal Air Force, Keir Giles spent the early 1990s working with aviation in the former USSR. He later joined the BBC Monitoring Service, specialising in economic and military coverage of Russia for UK government customers. From late 2005 Keir was attached to the UK Defence Academy's Conflict Studies Research Centre (CSRC), where he wrote and advised on Russian defence and security issues, and Russia's relations with its neighbours in Northern Europe. In 2010, Keir brought the CSRC team into the private sector to establish an independent consultancy.

Forrest Hare is an adjunct professor at Georgetown University and a Colonel in the United States Air Force. In his most recent assignment, he contributed to developing the United States Department of Defense cyberspace operations strategy. In addition, he has served in numerous information operations positions world-wide. Dr. Hare will be instructing on National Security Policy for Cyberspace at the GU Center for Peace and Security Studies. He has taught Economics and Geography at the United States Air Force Academy and the University of Maryland Asian Division. He received his Bachelor of Science degree from the United States Air Force

Academy, and a PhD from George Mason University. He also conducted post-graduate studies at the University of Fribourg, Switzerland, under a Swiss University Grant.

Kim Hartmann studied Computer Science and Mathematics at the Royal Institute of Technology, Stockholm, Sweden and at Otto-von-Guericke-University Magdeburg, Germany. She specialised in computer security, technical computer science and mathematical modeling. Kim Hartmann has worked on protocol security analysis, mathematical computer security modeling, computer security risk assessment and risk analysis of critical network infrastructures. Her research interests are secure network design principles, risk analysis and assessment of networks, network components and protocols. Since October 2011 Kim Hartmann has been employed at the Institute of Electronics, Signal Processing and Communication at Otto-von-Guericke University Magdeburg, Germany.

Jason Healey is the director of the Cyber Statecraft Initiative of the Atlantic Council, focusing on international cooperation, competition and conflict in cyberspace. He also is a board member (and former executive director) of the Cyber Conflict Studies Association and lecturer in cyber policy at Georgetown University. He is the principal investigator for the first book on cyber conflict history and his ideas on cyber topics have been widely published, including by the National Research Council, academic journals such as those from Brown and Georgetown Universities and the Aspen Strategy Group and think tanks.

Professor Dr. **Wolff Heintschel von Heinegg** is a Professor of Public Law, especially public international law, European law and foreign constitutional law at the Europa-Universität Viadrina in Frankfurt (Oder), Germany. He was among a group of international lawyers and naval experts who produced the San Remo Manual on International Law Applicable to Armed Conflicts at Sea. He is a member of several groups of experts working on the current state and progressive development of international humanitarian law, including the Manual on Air and Missile Warfare (2010) and the Manual on the International Law Applicable to Cyber Warfare.

Harsha K Kalutarage is a PhD student at the Digital Security and Forensics (SaFe) Research Group at Coventry University, UK. The SaFe group is involved in research for security and protection for scalable systems and critical infrastructure, with emphasis on transport, agriculture and unmanned systems. Harsha is currently also involved in a national project exploring efficient and reliable transportation of consignments (ERTOC) to help the transport and logistics sector design and operate a real-time carbon tracking system for logistics operations. The project is funded by the Technology Strategy Board (TSB) and the UK Engineering and Physical Science Research Council (EPSRC).

Assaf Keren is the product manager of Verint® Cyber Security division. Prior to joining Verint, Mr Keren held the post of Security Director for Israel e-Government. During his time in this office, Mr Keren was responsible for protecting Israel's government network, known to be one of the world's most attacked networks. He was responsible for a range of security issues ranging from application security, infrastructure security, methodologies, incident response and research. Following this post, Mr Keren served as a strategic consultant for large enterprises and organisations seeking to efficiently secure their complex network environments.

Robert Koch is an IT Staff Officer and Weapon Engineering Officer in the German Navy and a postdoctoral/external research associate at the Universität der Bundeswehr, where he is Chair for Communication Systems and Internet Services. He is led by Prof. Dr. Dreo Rodosek, part of the Munich Network Management Team. His key aspects of research activity are network security and intrusion detection. From 2008 to 2011, he was a research assistant at the Universität der Bundeswehr München, where he also studied Computer Science from 1999 until 2002. The topic of his doctoral thesis was Intrusion and Extrusion Detection in Encrypted Environments.

Professor **Samuel Liles** is an Associate Professor at the National Defense University in the Cyber Integration and Information Operations Department of the iCollege. Previously, Professor Liles was an Associate Professor at Purdue University Calumet. Professor Liles has written extensively on non-state actors, cyber warfare as a low-intensity conflict, conceptual analysis of cyber warfare and non-traditional assumptions of information assurance and security discipline. With over 30 years of experience and having served in the military, law enforcement, industry and academia, his current research interest is cyber conflict and forensics as a form of attribution. His PhD is from Purdue University.

Birgy Lorenz is a PhD student from Tallinn University, Estonia (syllabus Information Society Technologies). She is also eSafety trainer in Estonia in the Safer Internet in Estonia EE SIC Programme. Her activities include involvement in developing the National Curricula ICT syllabus, writing articles about e-safety and project management for the TurvaLan project. She has recently been awarded the Microsoft (2009) Innovative Teacher Award and the European Schoolnet (2010) first eLearning Award in the 'Internet Safety' category. She has also been crowned 'Teacher of the Year' in Estonia (2011). She is a member of the Cyber Defence Unit of the Estonian Defence League.

Hiro Onishi is a specialist at Alpine Electronics Research of America, Inc. He has been researching ITS (intelligent transportation system) and intelligent vehicle technologies for more than 25 years. He is currently a member of the TRB (Transportation Research Board) Cyber Security sub-committee, SAE (Society of Automotive Engineers) Safety & Human Factor committee, ITS America Safety committee and other public committees and task-forces. He holds a Master's degree in Electronics Engineering from Nagoya University (Nagoya, Japan).

CDR **Jean Paul Pierini** graduated in law at the University of Florence. Following two years of legal practice he joined the Navy in 1994 and served in several positions, mainly in the legal branch. Since 2008 he has served as the Legal Adviser to the Italian Fleet Command in Rome, dealing with operational issues involving maritime operations. Besides professional commitments, CDR Pierini published frequently in the form of legal reviews, articles and comments on criminal law, human rights, conflicts of law in criminal matters and judicial cooperation.

Daniel Plohmann studied Computer Science at the University of Bonn, Germany and received his diploma in 2009. Since 2010, he has been a PhD student at the University of Bonn, as well as a security researcher for the Cyber Defense Research Group at the Fraunhofer Research Institute for Communication, Information Processing and Ergonomics (FKIE) in Wachtberg,

Germany. His main research field is reverse engineering, with a focus on malware analysis and botnet mitigation.

Jody Prescott is a Senior Fellow at the West Point Center for the Rule of Law, and a former US Army Judge Advocate. He served in four different NATO headquarters (IFOR, SACT, JWC and ISAF), and was an Assistant Professor at the US Army Command & General Staff College and the US Military Academy. His research and writing focus on cyber conflict, the role of women in armed conflict, alternative energy and national security, and the leadership and ethical lessons of the Holocaust. Prescott now works for the US Department of Homeland Security in Vermont, USA.

Emily Reid has worked at the MITRE Corporation as an Information Security Engineer for three years, where she works on both cyber research and development for MITRE's government sponsors. She has worked on producing SCAP content, researching resilient products in MITRE's RAMBO lab, and most recently has focused on MITRE's international cyber efforts. She graduated from Tufts University with a BSc in Mathematics and minor in Computer Science, and is currently working towards an MSc in Computer Science with a Security concentration at Boston University.

Professor **Michael Schmitt** is Chairman of the International Law Department at the US Naval War College. He was previously the Chair of Public International Law at Durham University in the UK, and Dean of the George C. Marshall European Center for Security Studies in Germany. Professor Schmitt is presently serving as Director of the Project on the Manual of International Law Applicable to Cyber War.

Chris Spirito has spent the majority of his career focusing on information security and rapid prototyping of innovative solutions to meet time-critical customer needs. Chris leads the International Cybersecurity work programme for MITRE as well as supporting a counter unmanned systems experiment and global healthcare initiatives. Prior to returning to MITRE, Chris was the Director of Technology for Lux Research, a nanotechnology and physical science consultancy based in New York City. Chris also serves on the Advisory Board of WiRED International, whose mission is to provide medical and healthcare information, education and communications in developing and war-affected regions.

Mariarosaria Taddeo holds a Marie Curie Fellowship at the University of Hertfordshire, where she works on Informational Conflicts and their ethical implications. She is also affiliated to the Information Ethics Group (IEG) at the University of Oxford. Mariarosaria obtained a European PhD in Philosophy at the University of Padua. Her PhD thesis concerned the epistemic and ethical implications of the occurrences of e-Trust in distributed systems. She is the co-editor of the volume "The Ethics of Information Warfare", Philosophy of Engineering & Technology published by Springer Books (with L. Floridi, in press).

Lieutenant-Colonel **Patrice Tromparent** is currently serving in the Defense Policy Department of the Directorate for Strategic Affairs (French Ministry of Defense). In 1994, he graduated from the French Air Force Academy with a Master's degree in CIS. He also holds a Master's

degree in Computer Science (National Polytechnic School of Toulouse) and in Information Science (Aix Marseille University). As a career CIS engineer, LtCol Tromparent has served in the French CAOC and in the Airborne Electronic Warfare Squadron. He joined the Directorate for Strategic Affairs after graduating from the German Armed Forces General Command and Staff College in Hamburg.

Enn Tyugu has a Dr. Sci. degree in Computer Science from Leningrad Electrotechnical Institute, and has served as a Professor of Computer Science at Tallinn University of Technology and Professor of Software Engineering at the Royal Institute of Technology in Sweden. He is a member of the Estonian Academy of Sciences and of the IEEE Computer Society. He has written computer science books in Estonian, Russian and English. His present position is Leading Research Scientist at the Institute of Cybernetics of Tallinn University of Technology, and Scientist-Advisor at the Cooperative Cyber Defense Center of Excellence. His research interests are in intelligent software, simulation and cyber-security.

Sean Watts is an Associate Professor of Law at Creighton University Law School where he teaches Constitutional Law, Federal Courts, Federal Habeas Corpus, Law of War, International Criminal Law, and Military Law. He also serves as a Reserve Instructor at the United States Military Academy at West Point. His primary research interest is international legal regulation of emerging forms of warfare. He recently served as a defence team member in *Gotovina et al*. at the International Criminal Tribunal for Former Yugoslavia. Prior to teaching, Professor Watts served as an active-duty US Army officer for fifteen years, participating in legal and operational assignments.

Dr **Katharina Ziolkowski** is a member of the legal service of the German Ministry of Defence, currently serving at the NATO CCD COE. Since she joined the legal service, Dr Ziolkowski served as the legal advisor to the project 'InfoOp and CNO Capabilities', as legal advisor and military prosecutor for the Airborne Operations Division of the German Armed Forces and as legal advisor and law lecturer at the NATO School, the US Army's JAG School and the German General Command and Staff College. Dr Ziolkowski studied law and political science in Berlin and Barcelona and is admitted to the German bar. She holds a law degree from Freie Universität Berlin, an LL.M. in International Law from the University of New South Wales (Sydney) and a *magna cum laude* doctorate from Freie Universität Berlin, awarded by the Friedrich-Naumann-Foundation and the German Society for Military Law and the Law of War.