2011
3rd International Conference on Cyber Conflict

C. Czosseck, E. Tyugu, T. Wingfield (Eds.)

PROCEEDINGS

7-10 JUNE, 2011 TALLINN, ESTONIA

CCDCOE
NATO Cooperative Cyber Defence Centre of Excellence
Tallinn, Estonia

IEEE
Advancing Technology for Humanity

2011 3rd International Conference on Cyber Conflict (ICCC 2011)

IEEE Catalog Number: CFP1126N-PRT
ISBN 13 (print): 978-9949-9040-2-0
ISBN 13 (online): 978-9949-9040-3-7

Copyright and Reprint Permissions

Printed copies of this publication are available from:

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

and

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com

Cover Design: Jaakko Matsalu

Produced by IEEE eXpress Conference Publishing

For details on producing a conference proceedings and receiving an estimate, contact conferencepublishing@ieee.org or visit http://www.ieee.org/conferencepublishing

Foreword

Annual Conferences on Cyber Conflict, organised by the NATO Cooperative Cyber Defence Centre of Excellence (CCD COE) every summer in Tallinn, have become regular events bringing together international experts from all fields of cyber security.

The *2011 3rd International Conference on Cyber Conflict (ICCC 2011)* is the continuation of last year's *CCD COE Conference on Cyber Conflict.* Since NATO CCD COE is aiming to continuously improve the quality of its annual conferences, in 2011 we have selected IEEE as a technical co-sponsor of this event. Academic papers that have passed a strict double-blind peer review following the quality standards of IEEE conferences will be published in print and digitally by IEEE, and thus, increase ICCC's footprint in the academic world. Future Centre's conferences are planned to carry forward and extend this good relationship with IEEE.

To support CCD COE's mission of enhancing cooperation and information sharing between NATO, NATO nations and private and public players in the cyber domain, the *2011 3rd International Conference on Cyber Conflict* serves as a knowledge and network hub for technology experts, national security thinkers, lawyers interested in cyber conflicts and experts of other closely related areas.

In 2011 the conference focuses on defensive and offensive aspects of Cyber Forces, combining different views on cyber defence and operations in current threat environments. This is not limited to a military perspective, but it also covers legal, strategic and technical perspectives on equal grounds.

The different aspects of the cyber domain are discussed in two tracks: *1) Concepts, Strategy & Law* and *2) Technical Challenges &*

Solutions. The academic papers, combined with presentations delivered by distinguished world-class experts, are presented in respective tracks.

Issues to be tackled in the *Concepts, Strategy & Law* track include: legal and strategic aspects related to deterrence in cyberspace, involvement of conscripts in case of cyber conflicts and the implementation of cyber offensive capabilities in NATO, national aspects of Russia and Sweden with regards to their cyber forces, and a case of data leakage in deployed theatres presented.

The *Technical Challenges & Solutions* track covers a variety of different technical disciplines, reflecting the complexity of cyber defence in its technical implementation. The prospects of enhancing cyber defence capabilities by the use of Artificial Intelligence, and ideas for improved Early Warning and Intrusion Detection Systems are presented, to be supplemented by a proposal on how to preserve organisational privacy in intrusion detection log sharing. Efforts needed for acquiring and setting up botnets and on the other hand taking them down are compared and discussed, providing an insight in one of the major current threats. Additionally, a radically different way of fighting cyber adversaries – considering them being in a virtualized "game board" – is presented.

Many thanks to all the people around the globe who have been involved in organizing the *2011 3rd International Conference on Cyber Conflict:* external Programme Committee members, Track Chairs, all the volunteers and of course the lovely CCD COE staff who have with enormous efforts made this great event happen again.

Christian Czosseck and Enn Tyugu
NATO Cooperative Cyber Defence Centre of Excellence
Tallinn, Estonia
June 2011

ICCC 2011 Table of Contents

About the NATO CCD COE

The NATO Cooperative Cyber Defence Centre of Excellence (NATO CCD COE) is the 10th Centre of Excellence to gain full NATO accreditation by the North Atlantic Council.

The Centre's mission is to enhance capability, cooperation and information sharing among NATO, NATO nations and partners in cyber defence. Located in Tallinn, Estonia, the Centre is an international effort that currently includes Estonia, Germany, Hungary, Italy, Latvia, Lithuania, the Slovak Republic and Spain as Sponsoring Nations. The Centre is not part of NATO command nor is it funded by NATO budget. Instead it is directed, tasked and funded by the Steering Committee consisting of representatives of the above mentioned Sponsoring Nations.

The Centre has taken a NATO-oriented, interdisciplinary approach to its focus areas. The work of the Centre is based on extensive information exchange, co-operation with NATO and NATO states as well as academia and the private sector.

NATO CCD COE's key activities include organising and providing support to cyber defence exercises, organising conferences and workshops, delivering courses and trainings, conducting academic research on narrowly selected fields, working on legal and policy issues touching the cyber domain, and studying national cyber security strategies as well as wider strategic concepts and their applicability to the cyber domain.

ICCC 2011 Programme Committee

In 2011, about 40% of the content presented throughout the conference is based on academic papers, which are included in these proceedings. These submissions had passed a double-blind peer review process following the quality standards for IEEE sponsored conference and were finally approved by the international Program Committee. This year about 40% of all submissions were finally accepted.

To respect the interdisciplinary nature of cyber security and defence, this year's conference articulates itself in two parallel tracks, **Concepts, Strategy & Law**, and **Technical Challenges & Solutions**, which also reflects in the co-chaired Programme Committee.

Conference Organisers

Conference Chair:
Col. Ilmar Tamm
Director of Cooperative Cyber Defence
Centre of Excellence (CCD COE)

Publication Chair:
Capt. Christian Czosseck

Publicity:
Ms. Liisa Tallinn

Local Arrangement:
Mr. Raivo Terve and **Ms. Leelet Nellis**

Treasurer:
Ms. Piret Ilves

Program Committee:

Program Committee Co-chairs:

Prof. Thomas Wingfield,
George C. Marshall European Center
for Security Studies,
Germany

Prof. Enn Tyugu,
NATO Cooperative Cyber Defence Centre of Excellence and
Tallinn University of Technology,
Estonia

Track Chairs:

Lt.Col. Marco de Falco and **Mr. Kenneth Geers**
Technical Challenges & Solutions Track

Mr. Rain Ottis and **Ms. Eneken Tikk**
Concepts, Strategy and Law Track

Members of the Programme Committee:

Lt.Col. **Andrea Martorelli**	Italian Air Force
Dr. **Catharina Candolin**	Finnish Defence Staff
Cptn. **Christian Czosseck**	CCD COE and German Armed Forces
Dr. **Corrado Leita**	Symantec, Germany
Prof. **Dorothy Denning**	Naval Postgraduate School, USA
Prof. **Enn Tyugu**	CCD COE and Tallinn University of Technologies
Prof. **Eric Talbot Jensen**	Fordham University School of Law, USA
Dr. **G.W. Ray Davidson**	Purdue University Calumet, USA
Dr. **Gabriel Jakobson**	Altusys Corp, USA
Prof. **George Kostopoulos**	New York Institute of Technology, USA
Cptn. (Eng.) **Giuseppe Mendico**	Italian Air Force
Dr. **Jozef Vyskoč**	VaF Rovinka and Comenius Uni. Bratislava, Slovak Republic
Dr. **Juan Lopez, Jr.**	Air Force Institute of Technology, USA
Assoc. Prof. **Julie Ryan**	George Washington University, USA
Mr. **Kenneth Geers**	CCD COE and Tallinn University of Technologies
Dir. **Lars Nicander**	Swedish National Defence College
Lt.Col. **Marco de Falco**	CCD COE and Italian Air Force
Dr. **Marieke Klaver**	TNO Defence, Security and Safety, Netherlands
Prof. **Marta Beltrán**	Rey Juan Carlos University, Spain
Assoc. Prof. **Michael R. Grimaila**	Air Force Institute of Technology, USA
Dr. **Paul Leis**	SEB Bank Estonia
Dr. **Pavel Laskov**	University of Tübingen, Germany
Prof. **Peter Martini**	University of Bonn, Germany
Mr. **Rain Ottis**	CCD COE
Dr. **Risto Vaarandi**	CCD COE and SEB Estonia
Assoc. Prof. **Samuel Liles**	Purdue University Calumet, USA
Prof. **Sérgio Tenreiro de Magalhães**	Universidade Católica Portuguesa, Portugal
Asst. Prof. **Stefano Zanero**	Politecnico di Milano, Italy
Prof. **Thomas Wingfield**	G.C. Marshall Europ. Center for Security Studies, Germany
Mr. **Vítor Sá**	Universidade Católica Portuguesa, Portugal

The Program Committee and the NATO CCD COE want to thank the following reviewers for their additional review efforts and constructive remarks they provided:

Mr. Karlis Podins	Mr. Pablo Andreu Barasoain
Mr. Jim Chen	Mr. Spiros Vennis

3rd International Conference on Cyber Conflict Sponsors

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

European Union
Regional Development Fund

Investing in your future

Biographies of Contributors

Authors

Adil Aslan, currently Lieutenant Colonel and instructor at Turkish Air War College, Istanbul received a Bachelor of Science degree in *Electronic Engineering*, a Master's Degree in *Management of Education*, and a PhD in *Instructional Design and Technology* from the Old Dominion University, Virginia, US. He served at Turkish Air Force as an instructor pilot between 1995 and 2002 and as a staff officer at NATO between 2006 and 2009. He has participated to several multinational operations and projects.

Dmitri Alperovitch is VP of Threat Research at McAfee. He leads the company's research in Internet threat intelligence analysis and correlation, as well as development of in-the-cloud reputation services. With more than a decade of years of experience in the field of information security, he has significant experience working as a subject-matter expert with all levels of U.S. and International law enforcement on analysis, investigations and profiling of transnational organized criminal and nation-state cyber espionage activities. In early 2010, Dmitri led the global team that investigated the Operation Aurora attacks and gave that incident its name.

Daniel Bilar is an Assistant Professor of Computer Science at the University of New Orleans, Louisiana. Daniel is a recognized expert in US Federal Court in matters of computer security, programming, risk profiling and general forensics. He has degrees from Brown University (BA, Computer Science), Cornell University (MEng, Operations Research) and Dartmouth College (PhD, Engineering Sciences). His PhD thesis work ("Quantitative Risk Analysis of Computer Networks") addressed the problem of risk opacity of software in wired and wireless computer networks was provisionally patented by Dartmouth. He is the first author of a dozen peer-reviewed technical publications, and lectures widely nationally and internationally, most recently at a NATO venue on Cyber Warfare in Tallinn (Estonia), at Sandia National Labs in Albuquerque (NM), and at NAVY SPAWAR Information Ops & Unexplored Topics in a New Warfighting Domains.

Susan W. Brenner is NCR Distinguished Professor of Law and Technology at the University of Dayton School of Law in Dayton, Ohio. She has spoken at numerous events, including the Department of Homeland Security's Global Cyber Security Conference, and chaired a Working Group that helped to develop the Toolkit for Cybercrime Legislation for the International Telecommunications Union. In 2009 Oxford University Press published her book, *Cyber Threats: Emerging Fault Lines of the Nation-States.* In 2010, Praeger published her most recent book, *Cybercrime: Criminal Threats from Cyberspace.*

Hayretdin Bahşi is chief researcher in the Information Systems Security Group of TUBITAK BILGEM, Turkey. He obtained his BSc and MSc degrees from the Computer Engineering Department of Bilkent University and his PhD degree from the Computer Engineering Department of Sabancı University. His research interest includes data privacy, network security and computer forensics. He has participated in many penetration test infosec auditing and infosec consultancy projects.

Leo L. Clarke is General Counsel of Washington Federal, Inc., Seattle, Washington, USA. He has published extensively about the legal aspects of computer security and technology risk management since 1998. He is the author of over 30 scholarly and practical publications, and he has lectured at more than 50 professional and business conferences from Seattle to Dubai. He holds a B.A. in Economics with Distinction from Stanford University and a J.D. from UCLA Law School. He is a member of Phi Beta Kappa and the Order of the Coif.

Christian Czosseck is Scientist at the NATO Cooperative Cyber Defence Centre of Excellence in Tallinn, Estonia. Being a soldier in the German Armed Forces (Bundeswehr) for more than 12 years he held several Information Assurance positions in the German military. Christian graduated first in his class in computer science at the Universität der Bundeswehr in Munich and is currently PhD student at the Estonian Business School in Tallinn looking into national cyber security and botnet defence related issues.

Luc Dandurand, in his times as an officer in the Canadian Forces (CF), he led the CF's Network Vulnerability Analysis Team, supervising vulnerability assessments of military networks across Canada and in theatre. He founded the CF's Red Team, responsible for conducting computer network attacks against military networks to assess their security, improve the military's response to such attacks, and

demonstrate their impact. At the Communication Security Establishment, he led a team that developed novel solutions to difficult Cyber Defence problems. As a senior scientist at NC3A, he has contributed to scientific projects in Cyber Defence, focused on monitoring, information sharing and dynamic risk assessment.

Murat Dogrul received his B.S. degree in Electrical Engineering from the Turkish Air Force Academy, Istanbul in 2002. Followed by basic pilot and F-16 fighter training in Turkey, which he finished successfully in 2005, he was assigned to the 192nd Squadron at 9th Main Jet Base in Balikesir. In 2008 he graduated at the Air Force Institute of Technology, Ohio, USA, now holding a M.S. degree in Electrical Engineering. Currently in the rank of a Captain, he is student officer at the Turkish Air War College in Istanbul.

Victoria Ekstedt is the legal adviser for the Swedish Armed Forces Computer Network Operations Unit since 2007. She holds a Master of Law degree from Uppsala University, Sweden and a Master degree in Maritime law from the University of Southampton, England. Victoria has practised commercial law and served as legal adviser for the armed forces in Bosnia-Hercegovina 2004 and 2005. She is also a former military officer from the 1st Amphibious Regiment of the Swedish Armed Forces.

Keir Giles is a writer on defence and security issues affecting Russia. He is a director of the Conflict Studies Research Centre (CSRC), formerly the UK Ministry of Defence's centre for open-source research on Eurasian security and now an independent body. Keir is based in Oxford but travels widely briefing and presenting on Russian military affairs.

Mario Golling graduated in *Business Informatics* at the Universität der Bundeswehr München (University of the Federal Armed Forces Germany) in 2007. Till 2010, he was responsible for trainings and concept development in the fields of network administration, distributed simulation, operations research and IT security at the IT-branch of the "*Signals and Intelligence School of the Federal Armed Forces Germany*" (Führungsunterstützungsschule der Bundeswehr). Back at the Universität der Bundeswehr München, he is now research assistant at the continuative study program (*Studium plus*) at the *Chair for Communication Systems and Internet Services*. His research focus lies on network security, cyber defence, intrusion detection and next generation Internet.

Gabriel Klein is a research fellow at the Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE where he is investigating aspects of network security. His research interests include reactive security, security in MANETs and IT security common operational pictures. Gabriel studied computer science at the University of Bonn, Germany, with the main focus on computer networks and is now pursuing a doctoral degree.

Robert Koch is a research assistant at the Chair for Communication Systems and Internet Services led by Prof. Dr. Dreo Rodosek, part of the Munich Network Management Team. His research focus is with network security and intrusion detection. From 2005 to 2008, he was Deputy Weapon Engineering Officer and responsible for IT-Security on a German frigate. This included supervision of the ship's mainframe, networks and communication systems as well as the configuration management. After his military trainings at the Officers School, Operational School and Technical School of the German Navy he graduated in computer science at the Universität der Bundeswehr München in 2002.

Felix Leder is working as an innovation and new technologies architect for Norman ASA. After starting with Nokia he turned to his favourite field of research: IT-Security. During the time he worked for Fraunhofer and the University of Bonn, he joined into researching botnet mitigation tactics and new methodologies for executable and malware analysis. The results are several successful botnet takedowns, even using non-standard approaches. A lot of his spare-time is spent on involvement in the Honeynet Project and further proactive mitigation research.

Albert Levi received B.S., M.S. and Ph.D. degrees in Computer Engineering from Boğaziçi University, Istanbul, Turkey, in 1991, 1993 and 1999, respectively. After serving as visiting faculty member at the Oregon State University, USA till 2002, he has been faculty member of Computer Science and Engineering at the Sabanci University in Istanbul, Turkey where he is also founding co-director of the Cryptography and Information Security Group (CISec). He was promoted to associate professor in January 2008. His research interests include computer and network security with emphasis on mobile and wireless system security, public key infrastructures (PKI), privacy, and application layer security protocols. He is editorial board member of The Computer Journal published by Oxford University Press.

Fabio Mulazzani received his laurea degree in Strategic Science majors in Telecommunications (2004) from the University of Torino (Italy). He holds a Ph.D. in Computer Science from the Free University of Bolzano-Bozen (Italy). In the A.Y.s 2007-08 and 2008-09, he was Adjunct Professor of Information Processing, Software Metrics and Business Information Systems at the Faculty of Computer Science of the Free University of Bolzano-Bozen (Italy). In the A.Y. 2009-2010, he was Contract Professor of Business Information Systems at the Faculty of Computer Science at the Free University of Bolzano-Bozen (Italy). He works for the 2nd Signal Corps Alpine Regiment of the Italian Army in Bolzano.

Brendan Saltaformaggio is a junior studying Computer Science at the University of New Orleans. His research interests are Information Assurance, Mobile Device Security, and Operating Systems.

Salvatore Alessandro Sarcia' holds a Ph.D. in Informatics and Automation Engineering from the University of Rome *Tor Vergata* (Italy). From 2006 to 2008, he was at the Department of Computer Science at the University of Maryland (USA) as visiting researcher and is a NATO Defense College Senior Course 117 Ancien. His research interests are Computational Intelligence, Empirical Software Engineering, and Economics. Currently in the rank of a Lt.Col. he works at the Italian Army General Staff in Rome (Italy).

Björn Stelte is a research assistant at the Chair for Communication Systems and Internet Services led by Prof. Dr. Dreo Rodosek, part of the Munich Network Management Team. His current research focus lies on network security, intrusion detection and the development and application of secure Wireless Sensor Networks. He was involved in several studies for the European Commission, the Federal Armed Forces and the Federal Office for Information Security with focus on computer network and operating system security.

Enn Tyugu has Dr. Sci. degree in computer science from St. Petersburg Electrotechnical Institute, he has served as a professor of computer science at Tallinn University of Technology and at Royal Institute of Technology (KTH) in Stockholm. He is a member of the Estonian Academy of Sciences, of the Academia Europaea, of the IEEE Computer Society and of the Estonian IT Society. His present position is leading research scientist at the Institute of Cybernetics of Tallinn University of Technology and scientist at the Cooperative Cyber Defense Center of Excellence. His research interests are intelligent software and cyber-security.

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

Conscription and Cyber Conflict: Legal Issues

Susan W. Brenner
NCR Distinguished Professor
of Law & Technology
University of Dayton School of Law
Dayton, Ohio USA
Email: susanwbrenner@yahoo.com

Leo L. Clarke
R.O.I. Legal Group, Pllc
15 Ionia Ave SW, Suite 510
Grand Rapids, Michigan USA
Email: leo@roilegal.com

Abstract- **This paper examines legal issues that could arise from utilizing a civilian cyber defense corps to defend a nation-state and its assets from cyber attacks. We use Estonia's Cyber Defense League as an analytical device, and we examine issues that may arise under the CDL as it is currently configured and as it might be configured. Our analysis focuses on ten specific issues. We argue that the nature and inherent ambiguity of cyber war will require a reserve corps of IT specialists who can be conscripted if there is a substantial likelihood that a cyber attack will materially disrupt the public order. We also consider the practical and legal aspects of the criteria to be used to select conscripts and factors that will affect the duration of conscription.**

Of course, IT specialists do not work in isolation from the intellectual property and other IT assets owned by their private sector employers. The paper analyzes the issues raised by this symbiosis, including the risk that employers and other owners of assets will be treated as combatants by the cyber attacker, the potential legal issues created by the intellectual property rights of licensors, the potential unintended consequences affecting competition as conscripts defend a competitor of their private sector employer, and the privacy rights of third parties in data necessarily disclosed as part of defense activities. Finally, we consider whether the use of IT assets by conscripts entitles the asset-owners to compensation for the government's taking of their property and whether traditional notions of conscientious objection apply to cyber warfare.

Keywords: conscript, conscription, cyber conflict, cyber warfare, combatant, non-combatant, intellectual property, infringement, conscientious objection, kinetic warfare, Geneva Conventions

I. INTRODUCTION

In an article published in 2010, we analyzed the permissibility of conscripting civilians into a cyber war initiative under United States law [1]. Our premise was that conscription might be necessary if the government could not attract sufficient technological expertise to protect the public interest. Conscription, in other words, would allow the government to obtain the services of IT specialists who declined to assist in defending cyber conflicts because they determined they would be better off in private employment even if cyber attacks were successful.

In this article, we explore the legal and practical issues that are likely to arise when a country embarks on what we refer to as cyber conscription. Our analysis includes not only issues affecting the conscripts but also those who own the IT assets that conscripts would employ in the course of their duties.

We use Estonia's newly-created Cyber Defense League (CDL) as an analytical device [2]. We examine issues that may arise under the CDL as it is currently configured, and as it might be configured. Our analysis focuses on ten issues, each of which is examined below.

II. CONSCRIPTION FOR ATTACK-PREPARATENDESS OR DEFENSE

Since cyber attacks are inherently ambiguous in terms of source, intent, scope and duration, it is reasonable to assume that a cyber conscription program will be anticipatory, i.e., will be implemented before attacks occur or are expected. That brings us to the nature of the attacks: Based on what happened in Estonia in 2007 and in similar attacks, we believe it is reasonable to assume that cyber assaults will be of relatively limited duration, as opposed to the sustained assaults that have characterized kinetic warfare.

We based this assumption on several factors, one of which is that, unlike kinetic warriors, cyber attackers do not have to be physically present on the targeted state's territory; kinetic attacks tend to be prolonged because they are part of a zero-sum struggle to achieve a certain objective, e.g., gain control of territory, and because they are predicated on a mobilization of men and materiél. Cyber attackers operate remotely, and may have very different objectives; an attack, or series of attacks, may be the objective in and of itself. The attackers' goal may simply be to take targeted systems offline for some period of time, to demonstrate their ability to do so and/or the victim state's inability to prevent them from doing so, either of which could undermine the victim state's security.

For these and other reasons, we believe the appropriate model for cyber conscription is an as-needed force -- a version of the "National Guard" or "reserve" forces that are formed and trained before need arises and are "called up" to active

service when the need does arise [1]. As we note below, from what we know of the CDL, it seems to conform to this model.

III. WHEN SHOULD CONSCRIPTS BE ACTIVATED?

The inherent ambiguity of cyber conflict also raises the issue as to the appropriate criteria for activating conscripted reserves. The argument could be made that military forces should not be used for "mere law enforcement" because that is the role of local police. Under this argument, cyber conscripts should not be activated unless there is clear evidence of a nation-state sponsored attack. Unfortunately, experience shows that the actual sponsor and even the source of an attack will remain ambiguous long after the attack has ended. Therefore, a presumption against nation-state involvement would typically render use of conscripts ineffective.

We believe that a better analog is the use of national guards or reserve militia to enforce domestic laws in times of riot and other civil unrest. Activation in these circumstances is justified on the ground that the police force would not able to maintain public order without additional resources. Applying that approach to cyber defense might suggest that the conscript reservists should be activated if there is a substantial likelihood of a material disruption of the public order. For example, reservists would not be activated to defend attacks on non-essential services where the only potential losses are economic – such as an attack on large e-commerce sites, but would be activated where life or health were jeopardized – for example power grids, water supplies or medical facilities.

IV. CONSCRIPT SELECTION AND QUALIFICATIONS

When the U.S. Army drafted Elvis Presley in the 1950s, it was not for his singing voice and when it drafted Muhammad Ali in the 1960s, it was not for his boxing skills. Conscription has historically been a *levee en* masse or a lottery, not a targeted selection of individuals with specialized talents to perform particular functions. That aspect of conscription derives from the fact that until recently, massed manpower was the predominant engine of warfare.

The engines of cyber warfare are very different, which means the selection process must entail much more detail than the typical "draft registration" – name, age, address, education, physical condition and occupation. Conscripting IT personnel would require more detailed information about education and work experience, including familiarity with various platforms, software and industries. This means the selection process would require much more effort and planning on the part of the government and more response effort from the potential conscript.

Since complex IT functions generally require teams of professionals to coordinate their efforts, cyber defense would be most effective if entire "squadrons" were conscripted at the same time. Thus, conscription might be conducted by drafting

the workforce of a particular corporation or government agency. (Even government employees must be conscripted since they would otherwise be free to terminate their employment with the government and avoid service.) As we explain below, depending on how it is structured, such conscription could raise competitive and equitable considerations.

Similarly, the nature of cyber attacks will likely require a certain degree of specialization reflecting the IT structures and practices of specific industries. For example, IT specialists employed by financial institutions are unlikely to have the knowledge to respond to attacks on the electrical grid. Prompt and effective defense would, instead, require conscription of specialists who are responsible for designing and maintaining parts of the grid. Therefore, the conscription program may require more sophisticated organizational structures to ensure that specialized talents can be employed to their highest and best use as attacks affect different industries and locales. For example, command structures might have to adopt non-traditional approaches involving dual reporting according to both expertise and industry experience.

V. DURATION OF CONSCRIPTION

According to reports, the CDL is currently a voluntary "cadre of computer specialists" who will defend Estonia's computer infrastructure [2]. Since the CDL is part of the Defence League, we assume CDL members occupy a status analogous to that of members of the National Guard or reserve forces of other countries. That is, we assume CDL members can be called up to active military service, which in this context would involve cyber conflict.

If that is true, CDL members, like members of analogous units established in other countries, can presumably qualify as combatants under Article 4 of the Third Geneva Convention [3]. That is, members of a CDL-type cyber reserve force (i) will become combatants when they are called to duty and (ii) will otherwise occupy the status of civilian noncombatants [3], [4].

This dichotomous status generally proves unproblematic in the context of real-world warfare. In kinetic warfare, a reservist's status shifts from civilian to combatant when he is called to duty, and persists as long as he is on active duty with the military. The period of active duty is likely to last for weeks, months, even years. There is, therefore, a defined, temporally stable shift from one status to another; the clarity of this shift is enhanced by the fact that the reservist is usually summoned to serve his active duty in a location other than that where he lives as a civilian, is required to wear a uniform (versus civilian clothes) and engages in traditional martial activities.

Like conventional reservists, cyber conscripts will be called to active duty, but the nature and duration of that duty will differ from that of traditional reservists. Logically, there are two ways to structure the activation of cyber conscripts: One is

to activate them when a cyber attack is in progress or is imminent; in this alternative, the period of conscription would be coterminous with the length of the attack or the attack alert. Once the attack, or the threat of an attack, ended, the cyber conscript could be relieved of duty and return to civilian status.

The other option is to have cyber reservists permanently activated, on the not-unreasonable premise that they may need to respond to cyber attacks with little, if any, notice. This option effectively deprives cyber reservists of their civilian status, which we believe means it is neither a viable nor a necessary alternative. We do not see it as a viable alternative because it would presumably mean that members of a CDL-style cyber defense corps were full-time members of the military and, as such, unable to accept civilian employment. The countries that elected to implement this option would, therefore, deprive themselves of the services of an essential cadre of trained computer professionals. Countries bore this burden in other wars, such as World War I and World War II, because they had no other choice and because the conscription had an end point, i.e., draftees served "for the duration of the war" [5]. At this point, it does not appear that cyber conflict will have a determined end point, which means that this model of conscription could continue indefinitely.

We also do not see this model as a necessary alternative: Since activated cyber reservists presumably will not need to don a uniform, travel to a military base, or equip themselves with conventional weapons before they can participate in cyber defense, the situational activation option should be adequate.

Assuming that the reservist/activation model is adopted, the question arises as to how long the individual should be conscripted into reserve status. Given the costs of selection and training, there is a strong argument that conscription should be for a moderate length of time such as five years. Any longer period might be counter-productive because the rapid development of IT technology means that the skill sets required for effective defense will change rapidly. Therefore, the qualifications that led to conscription of a specific professional may not exist after five years, or the professional may have changed careers or specialties so that her skills are no longer needed.

It is also possible that conscription would not end on the expiration of a definite temporal period but on a conscript's termination of specified employment. This would be particularly likely if his conscription arose from his role with a particular employer or his involvement with particular IT assets. Since we presume that conscription is not cost-free to either the government or the conscript, there would be no value in continuing to train and include in defense planning, individuals who would not longer be of service. On the other hand, avoiding continued duty as a conscript should not be too easily achieved, since the rationale for the conscription program is that the government cannot rely on the voluntary cooperation of all individuals with requisite skills.

VI. SOLDIERS WITHOUT ARMS: THE NECESSITY FOR ACCESS TO PRIVATE IT TECHNOLOGY

One of the empirical distinctions between kinetic warfare and cyber warfare is the nature of the conflict: In kinetic warfare, confrontations between the two sides occur at a specific physical place; the forces of the respective parties engage in a struggle from which one side will emerge victorious. The struggle is conducted with conventional weapons, e.g., guns, tanks, explosives, provided by the warring states. The confrontations can, and do, occur on the territory of one of the states engaged in the conflict, but under the modern laws of war, the warring parties must make an effort to shield noncombatants from the struggle.

Cyber warfare is waged in cyber space but can wreak havoc in physical space by targeting components of a nation's critical infrastructure. The weapons used to wage cyber warfare differ from those used in conventional warfare in at least two ways: They do not involve the use of kinetic force; and they tend to be available to the civilian population. We assume that a CDL-style cyber defense force would not be composed of civilians with basic computer skills who would use their personal computer equipment to participate in cyber warfare.

We assume, instead, that because protecting a nation's critical infrastructure will be the primary objective in defensive cyber warfare, cyber defense forces will be composed of professionals employed by organizations that make up that infrastructure. In other words, we assume an embedded cyber defense force, one whose members can be called to active duty to defend the organizations for which they work. It seems reasonable to assume, therefore, that members of a CDL-style force will use the organization's IT systems to defend it.

That would suffice if we were analyzing a system that required infrastructure components to defend themselves, and only themselves, from cyber attacks. We, though, are analyzing a generalized cyber defense system, which presumably means that the employees of Infrastructure Component A would be authorized to use that entity's IT systems to defend it *and* other components of the nation's infrastructure. This generalized system could be executed in several ways. For example, the conscript could use his employer's IT assets to defend the IT according to military orders that differ from or supplement his employer's orders. Or, he could be ordered to defend the assets or business of a competitor of his employer, in effect providing benefits to the competitor at no cost to the competitor. Or he could be ordered to assist in defending unrelated assets because of his knowledge of specific technical issues or his general managerial and organizational skills.

This symbiosis between the human capital furnished by conscripts and the technology required for effective defense raises many complex issues, some of which we will discuss in the following sections.

VII. EFFECTS OF HOSTING CONSCRIPTS: POTENTIAL COMBATANT STATUS FOR INFRASTRUCTURE OWNERS

A fundamental issue is the status under international law the owners of IT infrastructure whose assets are used by cyber defense corps members in responding to attacks. It is likely that most of the individuals or companies that fall into this category will be the conscript's employer. As we saw in § IV, activated members of a CDL-style cyber corps will be combatants under the laws of war. The issue we take up here is whether the same is true of their employers.

Under Additional Protocol I of the Geneva Conventions, civilians lose their non-combatant status "for such time as they take a direct part in hostilities" [4]. Interpretative guidance for this provision says direct participation consists of "specific acts carried out by individuals as part of the conduct of hostilities" between warring states [6]. To qualify, such acts must (i) be likely to adversely the military operations of a party to the conflict or to injury persons or property, (ii) have a direct causal link with the adverse effect or injury and (iii) be specifically designed to cause the effect or injury [6]. Merely producing war materiél does not constitute direct participation, but a conscript's use of her employer's IT assets to defend an attack could be interpreted as not mere production of a weapon, but actual use of the asset as a weapon, even though it is used solely in defense [7].

In other words, a conscript's use of her employer's equipment or intellectual property in the course of carrying out her military orders could cause the employer to become a combatant and therefore a legitimate target for attack. This latter point would be academic if the employer is already under attack, but could present important issues when a conscript uses the employer's assets to defend another entity. And the argument that the employer's role constitutes direct participation in hostilities might be inferentially strengthened by the fact that the employer's authorization to the conscript encompasses the repeated use of the equipment for military purposes.

Another issue that might arise is whether use of an organization's computers might transform non-cyber corps employees who supported the efforts to repel the attack into combatants, on the same premise outlined above.

It is also possible that CDL-style cyber corps members could be activated to defend entities other than their own employers. Logically, this could occur in either of three ways: The CDL members could travel to another site to launch their defensive efforts; they could use their employer's computers to do so; or they could use computers that were in/near their employer's premises but reserved for cyber corps defense activities. The first scenario does not seem practicable if the need to respond is immediate; and if the attacks were part of a sustained series of attacks, this also might not be a viable option. Utilization of the second scenario would presumably raise the issue outlined above, with the additional factor that allowing

use of one's property as a weapon to defend another's property presents an even stronger case for finding combatant status. The third option, thought, would protect the employer from combatant status because using cyber corps-dedicated weapons would not implicate the CDL member's employer in the defense of a third party.

Logically, the "direct part in hostilities" issue could arise for another participant in any cyber war effort: the Internet Service Providers (ISPs). Since we are postulating a civilian-staffed cyber corps the efforts of which are primarily dedicated to defending civilian entities from cyber attacks, it is reasonable to assume that cyber attacks and the cyber corps' responses to attacks will all travel via commercial ISPs. The ISPs' role could, at least arguably, be construed as taking a "direct part" in the cyber hostilities; some have analogized the ISPs' role as the equivalent of using military aircraft to bomb enemy targets.

We are not asserting that the employers and co-workers of entities who employ members of a CDL-style cyber corps categorically become combatants by playing the roles outlined above. Nor are we making a similar assertion for ISPs whose systems carry defensive (and offensive) cyber attack signals. We simply note that the issue can arise in this context, which might make it prudent for a country developing a cyber corps to incorporate that possibility into its planning.

VIII. ECONOMIC CONSIDERATIONS: CONSCRIPT USE OF INTELLECTUAL PROPERTY

The symbiotic relationship between conscript and infrastructure creates another distinction between conscription for cyber defense compared to defense of kinetic warfare. The only requirements for someone drafted into traditional, kinetic military service are that the inductee be healthy, reasonably intelligent and not suffering from a mental disorder. The inductee's particular expertise – if any – is generally irrelevant (though it may play a role in his eventual unit assignment). The government provides all necessary lodging, food, equipment and training necessary to fulfill the conscript's obligations. The cyber conscript, in contrast, is drafted for his or her ability to bring specialized knowledge to bear in supporting the nation's sovereign integrity. That knowledge is likely to include information and ideas that are protected under intellectual property laws.

(We use "intellectual property" in its broadest definition to mean ideas, expressions of ideas and know-how including trade secrets, other proprietary information. These issues are made more complicated by legal doctrines that require an owner of intellectual property to take appropriate action to enforce its property rights at the risk of losing them against other parties.)

A conscript's access to intellectual property can become an issue even if he or she merely uses IT assets owned by the government. Assume, for example, that a conscript's executing orders requires her to use her knowledge of source code or

other proprietary information associated with third party software her employer had licensed. Her employer (or the third-party licensor) could seek to bar the conscript's carrying out her orders on the basis that she would necessarily use its intellectual property in doing so. The argument would be that such a use constitutes an infringement of the owner's property rights.

Rather than simply using licensed software, it is more likely that to carry out her orders, the conscript would need to revise or add code to a copyrighted software program licensed by her employer or another attack target. Such an act would probably constitute infringement, absent an appropriate license. And if the conscript's orders required her to access computers beyond the authority given by her employer, she might well be guilty of a criminal offense.

In short, absent a legislative solution, executing her orders could expose the conscript and the government to liability under intellectual property laws and/or under laws making it a crime to access a computer without being authorized to do so or in excess of one's authorized access. Therefore, in developing conscription legislation, consideration should be given to including a provision that addresses addressing a conscript's authority to use intellectual property licensed by her employer and/or others without paying a fee. The conscription legislation might, for example, grant the government a free, non-exclusive license to use all intellectual property that might be inevitably disclosed in the course of a conscript's service. (Whether this statutory license would constitute a taking is discussed in part X below.) Otherwise, cyber defense could be impeded by uncertainty and even litigation regarding conscripts' rights to use their employment-acquired knowledge in support of the defense effort.

IX. POTENTIAL UNINTENDED ECONOMIC CONSEQUENCES

The use of conscripts and related IT assets might also have unintended economic consequences.

Using conscripted forces to defend against cyber attacks raises one such issue because prudent management practices require governments and businesses to protect their IT assets and data even in the absence of a cyber war threat [8]. In comparison, kinetic warfare typically presents risks fundamentally different from those presented by "business as usual."

For example, the military's use of conscripted soldiers to defend a warehouse from invading forces inures to the benefit of the merchant owner as well as to the public at large. The warehouse is saved from destruction at no cost to merchant owner or his insurers. In those situations, however, the military effort indisputably arises from a risk not incurred in normal business circumstances – an unambiguously hostile attack by a foreign nation-state. Conscripts, on the other hand, may be ordered to defend against an attack that is not "military" in origin, but is "merely" cyber crime or cyber terrorism. (This risk, of course, arises from the inherent

ambiguity of cyber attacks.) The conscript might, therefore, be involved in implementing an IT defense that does not differ materially from defense against cyber crime. In this context, adoption of a conscription program might cause owners of IT to under-invest in IT security.

A different, but perhaps more significant, unintended consequence of conscripting corporate employees to defend cyber attacks is the potential conflict that might arise if conscripted employees of one organization were given access to another organization's IT assets or data to defend an attack. Since most IT systems are exceedingly complex and proprietary in nature, it is only reasonable to expect that the conscripts would have to work with employees of the target organization, and would have access to proprietary information concerning the target's customers, suppliers and other vendors. For example, assume that conscripted employees of Bank A were ordered to assist in the defense of an attack on Bank B, a competitor, and to mitigate resulting damage to the financial system. In the course of executing their orders, the conscripts might (i) disclose information Bank A had acquired at great cost to employees of Bank B, (ii) learn about strengths and weaknesses of Bank B's systems, (ii) be exposed to confidential pricing and other information granted to Bank B by vendors to both banks, and (iv) receive access to financial information of Bank B's customers which is protected by privacy laws.

Such a situation would not be acceptable to any of the affected parties. Bank A would not appreciate its competitor's receiving the benefit of its investment. Bank B would complain about the disclosure of its proprietary information. The vendors would allege a breach of confidentiality rights, and customers would allege breach of privacy laws. None of these consequences is a necessary result of the cyber attack; each is a real and likely substantial cost; and collectively the resulting harm may exceed that of the attack itself.

One response to these unintended consequences may be to preclude conscripts from communicating directly with competitors and instead require screening procedures. However, screening and similar procedures that required the insertion of third parties would introduce additional levels of complexity, delay and expense into situations that require immediate and efficient response.

X. COMPENSATION

As we explain below, compensation issues arise both for conscripted individuals and for third parties involved in a cyber conflict event.

Conscripts generally receive compensation from the military and forego the income from their pre-conscription private employment; this is considered to be a cost of citizenship. Reservists are typically paid at military scales while activated, although some employers may continue to supplement their compensation as a form of social responsibility.

There is, of course, a risk that an employer will terminate a conscripted employee because its business needs will continue to require services even if an attack occurs. Termination in this context is unlikely, however, because of the shortage of skills in the market place, the probable short activation period, the training and other transaction costs involved, and the likelihood that the replacement would also be conscripted. In light of these factors, it would not appear that either efficiency or equity would require special compensation provisions for conscripts.

The discussion above noted several situations in which a conscript's performance of his duties may cause his employer and/or third parties to incur costs or lose the benefits of bargains. These losses would result from the practical relationship between conscripts and IT technology typically owned by employers and those other entities. Given those financial consequences, owners of IT assets used by conscripts or infringed upon in the execution of orders might seek compensation from the government.

Estonia's Constitution, like the constitutions of many other countries, prohibits the government from taking private property unless the taking is in the public interest and for fair and immediate compensation. The taking or destruction of property in the course of warfare, however, is generally not considered to be a "taking" for such constitutional purposes [8]. Moreover, an asset-owner would not seem to have an equitable claim for compensation when the conscript is defending the owner itself.

On the other hand, the complexities of intellectual property law and the technology involved may counsel, as suggested above, including in the conscription legislation, an explicit grant of non-exclusive licenses to the government either at no cost or at a cost to be determined after the use is completed. Such explicit treatment would tend to reduce doubt, confusion and litigation and set the framework for consensual resolution of the appropriate amount of compensation that the government should pay for its requisition of assets for the war effort.

XI. CONSCIENTIOUS OBJECTION

Like many other countries, both Estonia and the United States recognize the right to refuse to serve in the military "for religious or ethical reasons" [9], [10]. If these or other countries decide to implement a conscript-style cyber corps, the issue of conscientious objection may arise. Since a cyber corps conscripts civilians into military service, the basic legal premise for conscientious objection seems to be established in this context. An issue may arise, however, as to whether conscientious objection is appropriate in conscription for cyber warfare.

Historically, conscientious objection was primarily based on religious or philosophical objections to the "obligation to use lethal force" [11]. While it is certainly possible that cyber attacks could result in a loss of life, the nature of cyber

combat is notably less lethal than kinetic warfare. This might, or might not, result in a lesser incidence of conscientious objection in this context. It also might, or might not, require countries to determine how traditional principles governing conscientious objection apply to cyber warfare.

XII. CONCLUSION

Our purpose in writing this paper is to identify many – but undoubtedly not all – of the legal issues that are likely to arise when a country elects to implement a CDL-style civilian cyber defense corps. Certain of the issues that will arise in a particular instance will, at least to some extent, be specific to the laws of that nation-state. Based on our research, though, we believe many of the issues are likely to be consistent, at least in countries that clearly demarcate civilian and military spheres of operation. It may be possible to address the more generic issues with international agreements or, perhaps, a template of model laws similar to the Toolkit for Cyber crime Legislation developed by the United Nation's International Telecommunication Union [12].

REFERENCES

[1] S.W. Brenner & L.L. Clarke, "Civilians in Cyber warfare: Conscripts," in Vanderbilt Journal of Transnational Law, vol. 43, (4), 1011, 2010.

[2] H. Kenyon, "Volunteer Cyber Corps to Defend Estonia in Wartime," Defense Systems, January 12, 2011.

[3] Geneva Convention (III) Relative to the Treatment of Prisoners of War, Aug. 12, 1949, 75 U.N.T.S. 135.

[4] Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I) Article 50, June 8, 1977, 1125 U.N.T.S. 3.

[5] Ex parte Billings, 46 F.Supp. 663 (U.S. District Court for the District of Kansas 1942).

[6] Assembly of the International Committee of the Red Cross, Interpretative Guidance on the Notion of Direct Participation in Hostilities 995 (February 26, 2009).

[7] Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949, 619 (Yves Sandoz et al. eds. 1987).

[8] S.W. Brenner & L.L. Clarke, "Civilians in Cyber warfare: Casualties," in Southern Methodist University Science & Technology Law Review, vol. XIII, (2), 2010.

[9] Constitution of the Republic of Estonia, Article 124(2).

[10] Welsh v. United States, 398 U.S. 333 (U.S. Supreme Court 1970).

[11] Special Rapporteur on Freedom of Religion or Belief, Framework for Communications: Conscientious Objection, Office of the United Nations High Commissioner for Human Rights.

[12] United Nations, International Telecommunications Union, ITU Toolkit for Cyber crime Legislation, 2010, http://www.itu.int/ITU-D/cyb/cyber security/projects/cyber law.html.

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

Cyber Security on Military Deployed Networks

A Case Study on Real Information Leakage

Cpt. Fabio MULAZZANI, Ph.D.
2nd Signal Alpine Regiment, Italian Army
and
Free University of Bozen/Bolzano
Bolzano, Italy
fmulazzani@unibz.it

Lt.Col. Salvatore A. SARCIA', Ph.D.
General Staff, Italian Army
and
University of Rome "Tor Vergata"
Rome, Italy
asarcia@disp.uniroma2.it

Abstract- **This paper reports on real information leakage occurred in a multinational mission. To investigate the nature of the leakage, we performed a survey among the military operators which showed that technical and cultural problems were key elements of the security shortfall. We also show that military deployed networks present some peculiarities with respect to infrastructure homeland networks. Therefore, the former should be managed differently from the latter. In particular, we highlight two reasons concerning either the operators or the networks: (1) Temporary nature of deployed networks and (2) Lack of training and guidance (es. SOPs). Finally, we propose a new approach that would strengthen the defense attitude of signal units and check whether protection activities are effective and reliable.**

Keywords: Military Field Data Network; Security Leak; Country Case Study; Human Resource Management, Cyber Offense

Disclaimer: This paper is a product of the authors designed to provide an independent point of view. It does not represent the opinions or official position of the Italian Army.

I. INTRODUCTION

On November the 28th 2010 several thousand classified documents were published on the Wikileaks website. This fact triggered cyber defense actions for almost all Governments targeted by the information leakage. How did this leakage happen? Firstly, personal responsibilities should have been taken into account, i.e. the soldier who stole the document and gave it to Wikileaks. Secondly, under what circumstances can highly classified documents be retrieved so easily? From a security perspective, governments' data networks (and, as such, those belonging to Armed Forces) supporting the treatment of classified files are usually properly supervised by a team of highly specialized and reliable servants defending the network from cyber threats. Such cyber defense services can then be easily maintained within national networks such as those within the national territory. However, military networks are present in a variety of areas of operations around the world, and they need to be managed (including security issues) in the location where they actually are. The administration of the network is locally-based because the link to the national data network is usually limited (e.g., at most 1 Mbps) – these circumstances make operational military networks quite different from the other governmental networks. It is clear that cyber-attacks would target the weakest area of a network, as well as operational networks may be considered as the target of cyber-attacks word-wide.

This paper presents a case study of a real incident occurred within an operational contingent that one of the authors dealt with some months ago. Moreover, it presents the analysis of the possible root causes that resulted in the incident. We compare and contrast domestic and operational (abroad) networks one another, trying to find the elements that make the military network deployed on the field so interesting for cyber security. The case study will describe the physical layer of the network and expand on software applications that determined security leakage. Our research question (closely stated below) is to figure out whether or not cultural factors such as background, educational level, and country-wide habits of the persons in charge of the network management (stakeholders) can be considered primary reasons of the security leak which we are referring to.

In order to investigate our research question, we performed a survey within the community of stakeholders involved in the management of the targeted network. The survey was oriented to (i) identify relevant security aspects and (ii) assess the stakeholders' awareness on cyber defense. Finally, from the analysis of the data collected from the survey we developed some proposals to address the issue of the security of military networks on field mission.

The rest of the paper is structured as follows. In section 2 we offer a review of the main IT security guidelines adopted by the Italian Army. In section 3 we propose a detailed description of the case study and the structure of the survey. In section 4 we illustrate the results of the survey previously described. In section 5 we illustrate our proposal for a new organizational approach for cyber defense. Finally, in sections 6, 7 and 8 we state the conclusions, the future works and the remarks.

II. RELATED WORKS

With [1] and [2] the Italian Army produced the directives on the security of classified and unclassified telecommunications and information systems, also on the basis of [3]. The purpose of the mentioned directives is to (i) clearly identify and define the organizational structure of the bodies responsible for the Information Security (INFOSEC) aspects; (ii) describe the formal procedure to require the homologation of the network; (iii) remind that operating systems must be certified according to international criteria like ITSEC or ISO/IEC 15408 – Common Criteria (CC) [4] [5] [6].

ITSEC is the acronym for Information Technology Security Evaluation Criteria and is a structured set of criteria for evaluating computer security. The evaluation consists in the examination of IT features and in a penetration testing of the Target of Evaluation (TOE). ITSEC identifies seven ascending levels of confidence that can be placed in the TOE, the levels are coded from E0 to E6. ITSEC can be seen as the natural evolution of the Trusted Computer System Evaluation Criteria, frequently referred to as the Orange Book [7] [8]. The Orange Book was commonly perceived as "too strict" in formal definitions that is because ITSEC has the purpose to create an environment flexible enough to identify new requirements sets when new security problems are found. In ITSEC is very important the concept of IS security requirements *reliability*. In particular, the *reliability* is seen as trust both in the *effectiveness* and in the *propriety* of the security systems that were designed and implemented. *Effectiveness* describes how the system responds to the attacks, and *propriety* identifies all the aspects related to the realization of the product.

ISO/IEC 15408 (CC) is a standard that aims to evaluate whether security facilities of information systems are properly designed and implemented. The CC supports understanding of "what the product does" (security functionality) and "how sure you are of that" (security assurance). From a practical perspective, CC provides a methodology, notation, and syntax to specify security requirements by means of three documents (Part 1, Part 2, and Part 3).

The CC aims at being a keystone for ISs *consumers*, *developers*, and *evaluators*. The CC states that any security analysis should examine the physical environment a system will exist in, the asset requiring protection, and the purpose of a system to be evaluated (target system). It then mandates a listing of the assumption, threats and organizational security policies, leading to a set of security objectives to be met. Using the objectives, a set of security requirements should be generated. Requirements that recur in various systems and settings become the Protection Profile (PP), which is intended to be reusable and defines the target system's security requirements known to be useful and effective in meeting the identified objectives, both for functions and assurance. The PP also contains the rationale for security objectives and security requirements. Evaluations, including various types of penetration testing, should then be carried out to determine a level of compliance with PP.

Even if ITSEC is being replaced by CC, a mapping between the two on the evaluation levels is given in [9].

A further support in the field of IT security is offered by the ISO/IEC 27000 family standard [10] that is a group of information security standards, it was developed on the basis of the publication BS 7799 "Code for Information Security Management" first issued in 1995 by the United Kingdom's Government Department of Trade and Industry (DTI) and the British Standard Institute.
ISO 27001 aims at offering a model for establishing, implementing, operating, monitoring, reviewing, maintaining, and improving an Information Security Management System (ISMS). ISO 27002 describes a set of information security management objectives and controls. ISO 27003 provides a set of guidelines to implement ISO 27000 standards. ISO 27004 provides a set of metrics to measure the efficiency of the ISMS. ISO 27005 provides a set of guidelines to conduct an information security risk management. ISO 27006 provides a set of guidelines to the various certification bodies on the process for certifying other organizations' ISMs. ISO 27007 provides a set of guidelines to those who audit ISMSs against ISO 27001, that indicate the best way to do so. Unfortunately, even if both public and private sector organizations have recognized the importance and benefits of ISO/IEC 27000 family, neither in [1], nor in [2], nor in [3] a reference to it was made. Further relevant research can be found in [11] and [12] where there are several theoretical and empirical studies that have been conducted with the purpose of offering models and frameworks aiming at better prioritizing cyber security threats.

III. CASE STUDY

The area of operations where the incident took place covered a wide area of the entire deployment territory. Among the common services such as telephone and radio network, the units were also served with three distinct data networks.
The first network, called "Lotus", was provided to the units by a Multinational Information Technology Service. Lotus was a VPN over Internet on which IBM Lotus software held about 50 clients. The purpose of Lotus was to offer (i) an Internet connection (especially in field of operations), and (ii) a support for collaborative tools such as the ones provided by IBM.
"Army-Net" and "Mission-Net" were managed by the signal unit. The former was a class "B" network by means of satellite links. Army-Net had about 400 users and was only employed by personnel of the contingent. Amy-Net supported several services such as: (i) Proxy internet connection; (ii) Unclassified information sharing (typically emails and documents); and (iii) Classified information sharing (usually preformatted messages) – the encryption of the signal was provided by ciphers connected to PCs allowing the treatment of classified information.
The third network, Mission-net, counted almost the same number of users as Army-Net, and it was settled to exchange both "Unclassified" and "Restricted" information; national classified information could not be shared over Mission-Net, however. The Mission-net access was provided to the units by means of microwaves backbones secured by Telesy KD03 IP cipher. Since the two potential security leaks were discovered in Mission-net, our case study focuses on this latter.

A. *MISSION-Net*

Before the discovery of the leaks, the Mission network had ten servers offering different services. They were as follows:

- Four Microsoft Windows 2003 Domain Controller Servers;
- One Web server based on Linux Debian;
- One Mail server base on Microsoft Windows 2003 and Altn Technologies MDeamon;
- One Microsoft Windows Server Update Services (WSUS) server;
- Two Sophos Anti-Virus servers working over a Microsoft Windows 2003 Server operating system;
- One FTP-Storage server based on Windows 2003 Server;

A military specialist, with the role of information system and network administrator, managed all services stated above. A military operator assisted the specialist in the daily work. The specialist was in charge of several duties, such as (i) repairing the network physically (ii) maintaining software applications, (iii) analyzing new process-oriented software applications not being supported yet; (iv) supporting the work of the Help Desk for the resolution of users' PC problems related to network services; (v) administering servers, routers, software licenses received for the mission.

B. *Relevant variables*

In order to investigate the nature of information leakage over deployed data networks of our case study, we identified some variables. Consequently, we devised some questions to survey some of those variables within an operation unit deployed in the field. The aim was to investigate to what extent the identified variables may be relevant to the explanation of the identified information leakage. Firstly, we identified the *turnover* of the operators as one of the variables to take into consideration. In fact, in the analyzed case, we noticed that the administrator changed with a frequency of a semester and, sometimes, of a quarter. The *inadequacy of a relevant percentage of users' PCs* was another identified variable. In the case study, about 30% of the users' PCs needed to be changed with more modern and adequate machines. The *length of the hand-over* from a contingent to another was considered as an additional variable. The incoming administrator worked only one week together with the outgoing administrator before the latter left. During that time, the incoming administrator received both the administrative and technical orders and hints to manage the systems. *Network topology change* was another variable relevant to the analysis. In the case study, the topology was also changed in terms of physical locations of servers. The *number of different locations* in which servers were dislocated seemed to be a relevant variable as well. In our case, the servers of Mission-net were distributed among four different locations. *Number of movements* was another relevant variable. We also identified as a relevant variable the *length of the relocation activity* as well as the *timeframe that the headquarters allotted* to the unit to complete the relocation.

C. *The Identified Leaks*

The first leak related to the possibility for those users not properly configured in the domain (i.e. with administrators rights) to access other users hard drive with the "C$" functionality. That functionality exists by default on Windows operating systems, and allows accessing other network users in anonymous mode – even without leaving any record on the access log file. With that functionality it is possible to have full permission (i.e., read, write, or delete) on the files of the remote hard drive. For example, a user that is not properly logged to the network domain, and has administrator rights, simply needs to write in the Windows Start menu run line the IP address of the PC being accessed followed by c$ (i.e. \\172.16.5.246\c$) and run the command.

The second leak related to the possibility, for those who could have accessed the mail server, to download all emails stored in the email server. Units and headquarters have not discovered yet whether or not someone downloaded information stored in the network servers. This is the reason why we refer to the information leakage as potential. In the program directory of MDeamon existed a folder containing a sub-folder for each hosted user. This folder hosted msg-format emails waiting for being downloaded by the local client. If either the administrator or someone else having the opportunity to access the mail server wanted to read the mail of a user, then the violation would be easy and painless. One simply would enter into the proper MDeamon folder and open the mail file using any editor. Additionally, reading attachments of a mail would not be problematic as well. Once the mail file would be opened with an editor, it would be necessary to copy and paste the attachment into an empty file with the proper extension.

D. *The Research Questions*

We identified two research questions:

a. *What technical and cultural factors affect security leaks within deployed data networks?*

b. *What actions are usually known by military operators for installing classified data sharing networks?*

   b.1. *What actions are usually known by those operators for increasing the level of security once a security leak is identified?*

Question a. aims at investigating technical (i.e. availability of technological devices) and cultural (i.e. security procedures knowledge) elements affecting the security of a deployed military network. It is worth noting that, signal units can usually be grouped into two categories: (i) those operating deployed networks and (ii) those operating non-deployed (i.e., infrastructure) networks. Questions b. aims at figuring out whether operators are prepared to install and operate secure deployed networks and (b.1.) whether those operators know the procedures to be taken after information leaks are discovered.

E. *Survey design*

Based on our experience in the field as specialists, we hypothesized that the problems stated above occurred for two reasons:

a) *Operators of deployed networks usually do not focus on security issues.*
b) *Network users are considered not to be harmful for the network security.*

We investigated these two hypotheses surveying those who operated the networks where information leakage was discovered. The survey we handed out is in annex "A". The survey is structured in six parts:

a) *Interviewed clustering. This part includes questions from 1 to 3 and aim to group the answers in homogeneous sets according to criteria like the rank of the interviewed, his background knowledge and his practical experience;*
b) *MISSION network knowledge/familiarity. This part includes questions form 4 to 8 and aim to collect (i) the intervieweed perceived security of that network, (ii) what are the features that they believe that contribute best to the security, (iii) the intervieweed knowledge on the MISSION network security features.*
c) *Question 9 is oriented to collect the perception of the intervieweed in creating or maintaing secured network like the MISSION.*
d) *Theoretical knowledge test. This part includes questions from 10 to 17 and are oriented to measure the theoretical knowledge (based on easy or medium diffiuculy questions) of the intervieweed on cyber security.*
e) *Question 18 is oriented to identify what is the perceived direction of a possible threat for the network.*
f) *Primary source of information.This part includes question 19 and 20 and aims to verify if the intervieweed (i) know what is the primary source of information for secured network and (ii) know where to get information for the procedures to adopt to install or maintain a secured network like mission.*

The questionnaire was submitted to 25 signal unit Officers, Warrant Officers, and Soldiers on duty in a Multinational mission where Italy participated in.

IV. DESCRIPTION OF THE RESULTS

The results of the survey are described below.
Interviewees clustering – (Question 1) the survey was submitted to a total of 7 Officers, 10 Warrant Officers, and 8 Soldiers. (Question 2) Almost all of the Officers with a rank not greater than Captain had a University degree in IS or Telecommunication topics, the rest of the people did not even have a high school diploma in IS or Telecom topics. The relevant thing was that 2 out of 3 Officers with rank greater or equal to Major did not have a degree (nor University or High School) related to IS or Telecom. (Question 3) All of the interviewees had a vast practical experience in the field because 14 of them participated in 2 to 4 missions, 5 of them 5 to 7, and the rest of them participated in 8 or more missions.

Mission network knowledge/familiarity – (Question 4) 17 interviewees perceive the Mission network as "secured enough", one Officer had "no idea", the rest of them were almost equally distributed between "very secured" and "somewhat secured", see Figure 1.

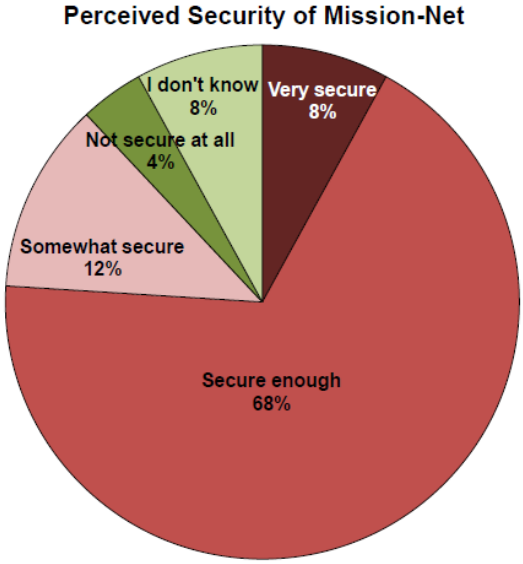

Figure 1. Answers to Question 4 – Perceived Security of Mission-Net

(Question 5) 19 interviewees believed that the sole contributor to the security of the Mission network was encryption, 4 argued that the security was provided by means of "network encryption" and "firewall", and 3 believed that a third contributor could be the anti-virus. Note that, no firewall system was installed in Mission-net. (Question 6) 18 interviewees knew that in the last year the Mission network was affected by viruses or trojans, 2 heard about "steeling information" or "unauthorized access" (this fact was reported by a soldier among the interviewees). The rest of the interviewees did not mention any significant event related to security. (Question 7) Almost all interviewees answered correctly. But, the answers were incomplete. They all knew that there was "network policies", but only few identified other factors as relevant to security. 2 interviewees maintained that a firewall was running to guarantee security. (Question 8) 18 interviewees did not remember whether a security check-up was ever performed, and 7 interviewees answered that the check-up was performed over the previous 6 month, see Figure 2.

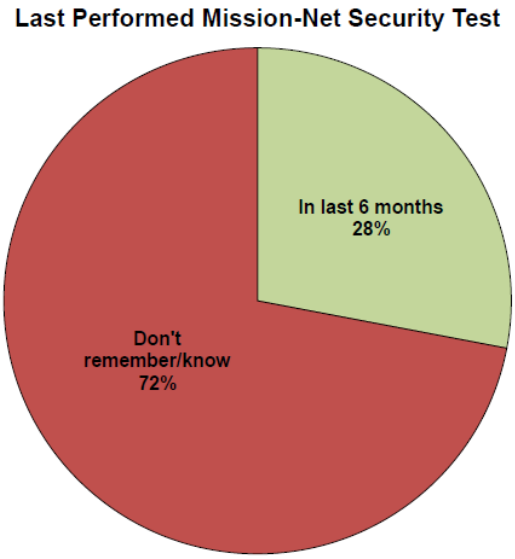

Figure 2. Answers to Question 5 – Last Performed Mission-Net Security Test

Interviewees' confidence in installing/maintaining a secured network – 17 interviewees believed to be "confident" in installing/maintaining a secured network. The rest interviewees selected "fair confident".

Theoretical knowledge test - The mean of the correct answer for all the interviewees was 45%. The mean for all the Officers was 54% of correct answers, Warrant Officers obtained 40% of correct answers, and soldiers 42%.

Perception of threat – (Question 18) the results showed that the great majority of the interviewees (20 interviewees) did not perceive internal users of the network as a threat for security.

Primary source of information – (Question 19) 19 Interviewees did not know the ISO standard proposed. Only one out of 6 asserted to know the standard. (Question 20) The interviewees stated that they used a Standard Operational Procedure (SOP) to install/maintain/supervise a secured data network as Mission-net. But 19 of them declared that the SOP was not provided by the line of command. 2 argued that the document was unavailable because restricted. 4 mistakenly declared that the document was available in a specified internal website.

From a general perspective, the survey showed that interviewees (i) believed that network security was primarily given by ciphers' physical encryption (ii) had poor knowledge of the primary concepts of cyber security (iii) did not perceive internal users of the network as a potential threat. We believe that this situation is due to two factors:

a) *Temporary nature of deployed networks,*
b) *Lack of training and guidance (es. SOPs).*

V. A NEW ORGANIZATIONAL APPROACH TO CYBER DEFENSE

Cyber defense is a relatively new area of concerns for governments and military alliances. The spread of technologies and the low cost of devises and machines turned an impressive number of people into potential information smugglers. Since there exists a rich market where stolen information can be traded, information sales have become a flourishing and profitable activity world-wide. Cyber-attacks aiming at worming out classified information generally take place through infrastructure networks of governments, companies, and institutions. Cyber-attacks against deployed networks are usually less frequent than the ones targeting infrastructure networks. However, the consequences of this kind of information leakage may be drastically severe for the troops deployed in the area of operations. Our case study shows that the fact that military operators underestimate potential information leaks is one of the main reasons for successful cyber-attacks against deployed networks. To assume that users of deployed networks are not able to attack deployed networks can eventually prove to be a great mistake. Secondly, before being deployed onto the field, signal troops should be trained on specific security aspects characterizing networks of interest.

Cyber-defense systems should be based upon three levels of defense. The first level should implement static protection such as identification, authorization, cryptographic protection, and access control. The second level should have mechanisms for collecting information and monitoring the state of the network. The third level should constantly evaluate the network protection [14]. Additionally, as our survey shows, operators' cultural aspects should be taken into account. To check whether or not the security level of our networks is effective we propose a new way of organizing cyber defense units (Figure 3).

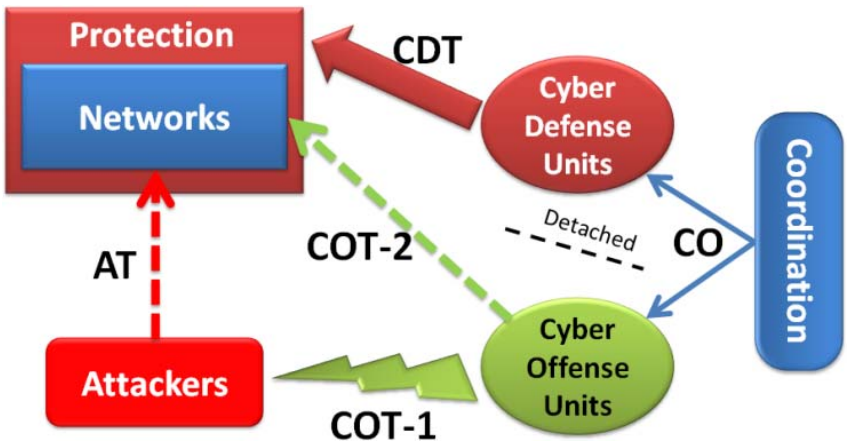

Figure 3. A new organizational approach to cyber defence.

Kotenko [13] proposes a multi-agent approach to cyber-security where teams of agents-malefactors, defense agents, and agents-users are simulated. However, Kotenko's approach cannot be applied to safeguard the security of our deployed troops because it is not always possible to have such simulating infrastructure for deployed networks. Moreover, we argue that a multi-agent approach is worth for experimenting and training signal units, but cannot guarantee the level of security required during operations. As showed in Figure 3, to safeguard our troops we propose a model which is still based upon three different bodies: defense, offense and malefactors (i.e., attackers) teams. However, we do not propose to delegate the assessment whether the level of cyber-security is adequate to a multi-agent framework. This assessment has to be done by a specialized team of people who can constantly evaluate the security and immediately report to the commandant of the mission.

The structure in Figure 3 can be used either for infrastructure or deployed networks. The novelty is that what we currently call cyber defense units should be split into two different kinds of units (detached): (i) those dealing with the protection (sheer defense) and (ii) those dealing with the offensive aspects of the defense. New cyber defense units should only have cyber defense tasks (CDTs) such as settling, maintaining, and protecting their networks. Cyber offense units should play two different roles: the role of attacker against external attackers – performing cyber offense tasks no. 1 (COT-1s) in Figure 3 – and the role of attacker against their own networks – performing cyber offense tasks no. 2 (COT-2s) in Figure 3 – To have detached units dealing with cyber offense only can better differentiate the preparation of offensive operators such that they can be focused on performing specific actions against external attackers. Employing a cyber-offense unit specialized in performing offensive tasks is worth for verifying whether or not the protection activity of cyber defense units is reliable and effective as expected. This situation would also strengthen the defensive attitude of cyber defense units since it would be 100% certain that either infrastructure or deployed networks would be under constant attack at least by cyber offensive units. In case of security shortfalls, the proposed organizational structure (Figure 3) would reduce the latent period between the beginning of the security problem and when the problem is discovered. This would reduce the probability that real attackers worm out sensitive information. In the case study, the operators realized the potential leaks after different months. On the other hand, a constant competition between cyber defense and cyber offense units would bring about an increase in security. Notice that, since cyber defense and offense units should operate synergically, a coordination function would be required. This would guarantee the integrity and the legal framework of the whole cyber activity.

The proposed organizational model stems from what, in science, is called "empirical approach". Empiricists emphasize those aspects that are related to evidence. Knowledge can only be discovered in experiments. We propose an empirical approach to cyber-defense meaning that the only way of assessing whether security of our information is guaranteed is to constantly try out this security through employing *ad hoc* offensive teams playing the role of attackers. Empiricism is the other side of the coin of *a priori* reasoning as, in statistics, prior

information is complementary to information arising from an empirical distribution. The idea of an empirical approach to cyber defense is to use both kinds of information: 1) *a priori* information, i.e. the one dealing with known and expectable attacks, and 2) empirical knowledge, i.e. what a real offensive team playing the role of attacker can do in the situation.
However, it is clear enough that instead of only having a multi-agent framework or a human-based security assessment approach, it would be better, when possible, if our troops could rely upon both systems.
It is important to note that cyber offensive units could only simulate attacks based on what is already known (*a priori* information). In other words, they cannot perform unknown attacks. Nevertheless, they may devise new attacks with the twofold aim of either increasing their ability or checking the network at a higher level of security (empirical approach). However, we know that the organizational structure depicted in Figure 3 is not enough to guarantee security of either infrastructure or deployed networks. Nevertheless, our proposal has a good potential for verifying and assessing the level of security that operators should guarantee on sensitive and sometimes critical information stored over the maintained networks.
Based on the results of the survey, we believe that it is mandatory that signal units develop SOPs to guide and standardize procedures of installation, maintenance, and administration of deployed networks. SOPs should also refer to authorized software applications. A good habit would be to describe information leakage in terms of lessons learned which would eventually update SOPs. What we really suggest is to change the way of thinking of cyber-defense in favor of an empirical approach. Only by trying out information security can we assess whether our defense system is effective.

VI. LIMITATIONS AND THREATS TO VALIDITY

It is always difficult to generalize data collected in only one environment. For this reason we argue that our analysis has some threats to external validity that have to be taken into consideration when using our conclusions in a more general context. However, our organizational solution can be applied without limitations either to infrastructure or deployed networks. Based on expertise of the authors in international environments, we maintain that problems discussed in the case study are commonplace. Consequently, even though the surveyed population is not representative of the statistical population of the signal operators, we believe that technical and cultural problems identified in our case study are fairly common to operators of other nations similar to Italy. Therefore, the proposed results are worthwhile and can be taken into account before the deployment of operational units.

VII. CONCLUSIONS

In this paper we described real information leakage which took place during a multinational operation where Italy participated in. Our research questions aimed

at investigating whether cultural and technical aspects concerning military operators could affect the security of deployed networks. Even though the limitation of the performed survey could not be completely generalized, we showed that deployed networks are settled based on the idea that they are temporary and then do not require high security measures. The second point was that military operators believe that cipher devises can solve all information leakage problems. We showed that this is not the case mainly because there is no guidance for those operators to avoid information leaks which are not dealt with by ciphers.
Finally we illustrated an organizational solution to cyber defense which we called "empirical approach to cyber defense" such that it would be better to have two different and detached kinds of signal units: (i) those dealing with installing, maintaining, and protecting networks (cyber defense units) and (ii) those dealing with offensive tasks against either real attackers or their own networks. This approach would strengthen the defense attitude of signal units and check whether protection activities are effective and reliable.

VIII. FUTURE WORK

It would be worth using the identified variables to statistically evaluate whether there exists a significant correlation between those (independent) variables and information leaks (binary dependent variable). We also argue that our empirical approach to cyber defense should be tested in field before being applied. Therefore, further research is required in order to investigate whether or not having two detached kinds of units (defensive and offensive) is worthwhile and viable in practical terms.

FINAL REMARKS

F. Mulazzani developed sections 1, 2, 3, and Annex "A"; S.A. SARCIA' developed sections 5, 6, 7. Sections 4, 8, 9 were developed jointly by the two authors.

ANNEX "A"

1. **What is your rank?** (a) Soldier (b) Warrant Officer (c) form 2nd Lt. to Cpt. (d) Major or above.
2. **Do you have a degree on Information Systems or Telecommunication issues?** (a) Yes (b) No; If yes, what is the degree level? (a)High School Diploma (b) BSc (c) MSc (d) Specialized Master Course.
3. **So far, how many missions (including the present) have you attended dealing with IS or telecommunications issues?** (a) 1 (b)2-4 (c) 5-7 (d) >8.
4. **How secure do you believe your network is?** (a) Very secure (b) Secure enough(c) Somewhat secure (d) Not secure at all (e) I do not know.
5. **Among the following aspects, which one do you consider the best contributors to the security of the Mission network?** – you can choose more than one - (a) network encryption (b) firewall (c) intrusion detection (d) identity (e) access control (f) traffic monitoring (g) vulnerability scanning (h) anti-viruses (i) other – specify.

6. **Have you had or known of these events in the Mission network in the last year?** (a) Viruses or trojan horses (b) Employees steeling information or allowing unauthorized access (c) Hackers targeting your systems (d) Lost or stolen backup tapes (e) Lost or stolen computers or data storage (f) None, if other specify.
7. **In the Mission network which of the following is allowed?** (a) Network use policies for employees (b) Automated patch management for security (c) Smart password policy (d) Spam control (e) Spyware protection (f) Virus protection (g) Firewall.
8. **When was the last time that Mission network was tested for security issues?** (a) More than one year ago (b) In last year (c) In last 6 months (d) In last 30 days (e) I do not remember or I do not know.
9. **How do you feel confident in either creating new or maintaining secured data networks like the Mission?** (a) Highly confident (b) Very confident (c) Confident (d) Fairly confident (e) No confident.
10. **We don't want our packets to get lost in transit. Which OSI layer is responsible for ordered delivery of packets?** (a) Network (b) Link (c) Transport (d) Physical
11. **What can a firewall protect against?** (a) Viruses (b) Unauthenticated interactive logins from the outside world (c) Connecting to and from the outside world (d) other.
12. **This is a program or file that is specifically developed for the purpose of doing harm:** (a) Buffer overflow (b) Bastion Host (c) Malware (d) Ping sweep.
13. **This is a program in which malicious or harmful code is contained inside apparently harmless programming data:** (a) War dialer (b) Spam trap (c) Trojan horse (d) email.
14. **A way of verifying a message's integrity after transport across a network is through the use of:** (a) A message authentication code (b) Steganography (c) An encryption key (d) A cipher.
15. **Which statement best describes the advantages of public key encryption**? (a) Keys are exchanged publicly without an eavesdropper being able to decrypt messages (b) Knowledge of one's public key does not yield knowledge of their private key (c) Encryption performance is faster than secret-key encryption (d) A and B only (e) B and C only.
16. **Which of the following best describes what is removed from a hard drive when a file is deleted from the hard drive?** (a) The MBR record, the FAT record, and the Directory Table entry (b) The FAT record, the Directory Table entry, and the data clusters that the file occupied (c) The FAT record and the Directory Table entry (d) The FAT record, the Directory Table Entry, and the Partition Table
17. **What is a secure process for keeping confidential information private?** (a) GnPG (b) PGP (c) network cipher (d) password protection (e) other.
18. **What do you think would be the main reason for most of the information security breaches?** (a) external hackers (b) poor programming (c) internal employees (d) bad firewall settings.

19. **Do you know ISO/IEC 15408?** (a) yes (b) No **– If yes what is that for?** (a) instructions to create secured security sytems (b) evaluation criteria for IT security techniques (c) cypher certification to be used in networks like Mission (d) other, specify.
20. **Do you use any SOP developed by the Army to install/maintain/supervise a secured network like the Mission?** (a) Yes (b) No – If yes, can you tell where this SOP is available (a) don't know, I got it from friends (b) it is a restricted document, can't tell (c) from an Army web site, please specify.

REFERENCES

[1] Department of the Army, Signal Soldier's Guide - Field Manual, March 2009.

[2] Italian Army General Staff – Security Office, "Software Systems, Telecommunication and Security – Unclassified documents, 2008.

[3] Italian Army General Staff – Security Office, "Software Systems, Telecommunication and Security – Classified documents, 2008.

[4] ISO/IEC 15408-1, Information Technology - Security techniques - Evaluation criteria for IT security - Part 1: Introduction and general model, 2009.

[5] ISO/IEC 15408-2, Information technology - Security techniques - Evaluation criteria for IT security - Part 2: Security functional components, 2008.

[6] ISO/IEC 15408-3, Information technology - Security techniques - Evaluation criteria for IT security - Part 3: Security assurance components, 2008.

[7] U.S. Department of Defence, Trusted Computer System Evaluation Criteria, December 1985, DoDD 5200.28-STD.

[8] U.S. Department of Defence, Directive: Information Assurance, October 2002, DoDD 8500.01 E.

[9] Bundesamt fuer Sicherheit in der Informationstechnik, "Application Notes and Interpretation of the Scheme (AIS): ITSEC to CC Mapping with Specific Attack Potential," 2010. [Online]. https://www.bsi.bund.de

[10] ISO/IEC 27000:2009 Information technology – Security techniques – Information security management systems – Overview and vocabulary (2009).

[11] F. Hare, "The Cyber Threat to National Security: Why can't we agree", in Conference on Cyber Conflicts, Tallin, Estonia, 2010, pp. 211-225.

[12] S. Liles, "Cyber Warfare: As a Form of Low-Intensity Conflict and Insurgency", in Conference on Cyber Conflicts, Tallin, Estonia, 2010, pp. 47-57.

[13] I.V. Kotenko, "Multi-agent Modeling and Simulation of Cyber-Attacks and Cyber-Defense for Homeland Security," IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Dortmund, Germany, 6-8 Sep. 2008.

[14] I.V. Kotenko, A.V. Ulanov, "Agent-based simulation of DDOS attacks and defense mechanisms," Journal of Computing, Vol.4, Issue 2, 2005.

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

Developing an International Cooperation on Cyber Defense and Deterrence against Cyber Terrorism

Murat Dogrul, Adil Aslan, Eyyup Celik
Turkish Air War College
Istanbul, Turkey

***Abstract*- Information Technology (IT) security is a growing concern for governments around the world. Cyber terrorism poses a direct threat to the security of the nations' critical infrastructures and ITs as a low-cost asymmetric warfare element. Most of these nations are aware of the vulnerability of the information technologies and the significance of protecting critical infrastructures. To counteract the threat of potentially disastrous cyber attacks, nations' policy makers are increasingly pondering on the use of deterrence strategies to supplement cyber defense. Nations create their own national policies and strategies which cover cyber security countermeasures including cyber defense and deterrence against cyber threats. But it is rather hard to cope with the threat by means of merely 'national' cyber defense policies and strategies, since the cyberspace spans worldwide and attack's origin can even be overseas. The term "cyber terrorism" is another source of controversy. An agreement on a common definition of cyber terrorism among the nations is needed. However, the international community has not been able to succeed in developing a commonly accepted comprehensive definition of "terrorism" itself.**

This paper evaluates the importance of building international cooperation on cyber defense and deterrence against cyber terrorism. It aims to improve and further existing contents and definitions of cyber terrorism; discusses the attractiveness of cyber attacks for terrorists and past experiences on cyber terrorism. It emphasizes establishing international legal measures and cooperation between nations against cyber terrorism in order to maintain the international stability and prosperity. In accordance with NATO's new strategic concept, it focuses on developing the member nations' ability to prevent, detect, defend against and recover from cyber attacks to enhance and coordinate national cyber defense capabilities. It provides necessary steps that have to be taken globally in order to counter cyber terrorism.

Keywords: cyber terrorism, terrorism, cyber defence, cyber deterrence

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the Turkish Air Force, Turkish Ministry of Defense, or the Turkish Government.

I. INTRODUCTION

The rapid evolution of information and communication technologies, and widespread services provided by the cyberspace bring up the question, "How can security of cyberspace be ensured?". IT and critical infrastructure networks are interconnected with each other, and can be accessed from anywhere in the world. In today's cyber world, a wide range of critical infrastructures from water supplies to transportation, from energy to communication technologies are vulnerable to cyber attacks. These infrastructures have little to none cyber protection, and rely on outdated conventional security solutions. A terrorist cyber attack on these industries could give rise to environmental disasters, economic casualties, and loss of property and/or loss of life. In this context, it is urgent that nations prepare for a possible cyber attack on critical infrastructure.

Plenty of investments have been made to prevent classical terrorist violence but the developed countries remain highly vulnerable to cyber attacks against the computer networks that are critical to national and economic security. The growing complexity and interconnectedness of these infrastructure systems, and their reliance on computers, not only make them more vulnerable to attack but also increase the potential scope of an attack's effects. This fear has prompted the governments to pump significant resources into protecting the critical national infrastructures [1].

In order to protect their vital interests, many technology dependent countries concentrate on organizing their cyber security policies. Most of these nations have taken some sort of national legal and military measures. But without international cooperation, these national measures are inadequate against cyber terrorism. Regional partnerships also do not provide adequate cyber security, since the cyber attacks can originate from off-region or off-partnership countries. In order to provide a worldwide international cooperation; the term "cyber terrorism" should be defined precisely and activities, considered as terrorist activity, should be determined as a first step. After that, developing both legislative and military collaborations should be discussed.

This paper first introduces the existing definitions and different aspects of cyber terrorism and relevant terms. It clarifies the extent of the cyber terrorism. Then it investigates how the terrorist organizations exploit cyberspace and why cyber domain is an attractive choice for terrorists. Next, it examines both legislative and military international cooperation attempts against cyber terrorism up to this day. Lastly, it offers an international game plan in order to set up defense and deterrence against cyber terrorism globally.

II. THE TERM "CYBER-TERRORISM"

While some authorities claim that there hasn't been any true cyber terrorism attack yet, others assert that terrorists already take advantage of the Internet. The source of this disagreement is inability to exactly define both "terrorism" and "cyber terrorism".

There is no universally accepted definition of terrorism and cyber terrorism; even when people agree on the rough definitions, they sometimes disagree about whether or not the definitions fit particular incidents. In order to understand terrorism, different views on what exactly constitutes terrorism must be assessed. Up to the present, no single international definition seems to satisfy the wide interpretation of terrorism or cyber terrorism.

However some authors were able to produce quite general definitions. In terms of its etymology, the word "terror" comes from the Latin word "terrere", meaning "to frighten, to terrorize, to intimidate"[2]. Usually, a series of terror incidents that are interconnected and directed at a certain political target is required in order to arrive a definition of terrorism. According to Bozdemir "Terrorism is a strategic approach which, for political purposes, identifies itself with a method which includes the use of organized, systematic and continuous terror."[3].

Denning defines terrorism as "The unlawful use or threatened use of force or violence by a person or an organized group against people or property with the intention of intimidating or coercing societies or governments, often for ideological or political reasons" [4].

Denning also defines cyber terrorism as; the convergence of terrorism and cyberspace. "It is generally understood to mean unlawful attacks and threats of attack against computers, networks, and the information stored therein when done to intimidate or coerce a government or its people in furtherance of political or social objectives. Further, to qualify as cyber terrorism, an attack should result in violence against persons or property, or at least cause enough harm to generate fear" [4].

The term cyber terrorism may be mixed up with "information warfare" and "cyber crime". But there is a major difference between cyber terrorism and information warfare. Cyber terrorism means premeditated, politically motivated attacks by sub national groups or clandestine agents, or individuals against information and computer systems, computer programs, and data that result in violence against non-combatant targets. But older term known as information warfare is defined as "a planned attack by nations or their agents against information and computer systems, computer programs, and data that result in enemy losses." [5].

Information warfare also encloses the term "cyber warfare". But cyber warfare's interest is limited to cyberspace. Information warfare and cyber warfare have "certain targets" in a war but cyber terrorism causes fear and harm to anyone in the targeted vicinity.

Along with these terms there is a phenomenon of cyber crime used frequently by law enforcement agencies. Although physical forms of cyber terrorism, information warfare, and cyber crime often sound very much alike, cyber crime is a crime committed through the use of information technology.

For instance; if a person hacks someone's banking account and/or steals credit card information, then it is called as cyber crime, because the attacker's intention is neither political nor social. If the same attack is directed to substantial number of banking accounts and the attacker declares that he is going to continue attacks until the government accepts his demands; moreover as a consequence of this attack people begin to fear and withdraw their money from the banks then it is labeled as

cyber terrorism. If the activities are carried out by the agents of a foreign power, and if all the banking system of a nation is targeted, then it could be labeled as cyber warfare. If the attacks to the banking system are not limited to cyberspace then it is called as information warfare.
Terrorist organization websites and the use of the Internet by the terrorists are other concerns. Most of the terrorists have not mastered the technology necessary for launching large scale attacks. However, some websites offer technologies for hire on the internet and provide information to reach bot-nets to execute "Distributed Denial of Service Attacks". Since the cyber terrorism is the convergence of terrorism and cyberspace, not only the devastating terrorist cyber attacks but also terrorist actions such as propaganda and recruiting carried out on the Internet should be considered as "cyber terrorism". Terrorist organization websites agitate public opinion, educate and motivate the members, command and control the organization, make propaganda to the target population and provide information to carry out cyber attack. Therefore, both the terrorist cyber attacks and the use of internet websites by the terrorists should be treated together and evaluated under the definition of the cyber terrorism.
After defining the extent of the term cyber terrorism, the exploitation of cyberspace by terrorists will be discussed.

III. WHY AND HOW THE TERRORIST ORGANIZATIONS EXPLOIT CYBERSPACE

There are many reasons that why cyberspace is an attractive choice for the terrorist purposes. Cyber attacks offer the capabilities for terrorist activities with wider-reaching impacts. Using cyber attacks, terrorists can inflict much wider damage to a country than they could by resorting to physical violence. With traditional terrorist activities, such as bombings, the impacts are isolated within specific physical locations and communities. Large part of the population acts only as observers and they are not directly affected by terrorist acts. The media and public attention is more likely to focus on the destruction of property and/or loss of life than whatever "cause" the activity was intended to promote. The ability of cyber terrorism activities to effect wider part of the population may give the groups involved greater leverage in terms of achieving their political and social objectives [6].
The motivation of the cyber terrorists comes from their political agenda. Their attacks are politically motivated and directed to specific critical system and infrastructures. This common agenda gathers all the hackers in the terrorist organization on the same goal. This collective action would do more harm than the action of individual hackers.
There are various reasons why cyber attacks are an attractive choice for terrorists such as;

- As terrorists have a limited amount of funds, cyber attacks are more tempting as they would require less people and less resources (meaning less funds). On the other hand, they can target and affect large numbers of

people with same amount of funds. In other words benefit to cost ratio is extremely high.

- It enables terrorists to remain unknown, as they could be far away from the physical location where the terrorism is being carried out. As terrorists normally set up camp in a country with a weak government, the cyber terrorist could set up anywhere and remain anonymous [7].
- Mostly, attacks are easy to carry out because many targets are poorly protected. Therefore attackers can choose from a wide variety of targets [8].
- When the attack is set up, it can be launched quickly without any need for further preparation [8].
- There are no physical barriers or check points that they have to cross [9].
- The speed and form of attacks are not dependent on the connection speed of the attacker. The connection speed of captured victim computers can be fully exploited [8].
- A combination of both physical terrorism and cyber terrorism is thought to be the most effective use of cyber terrorism [10].

In this regard, how the terrorist organization websites encourage the terrorist attacks should be revealed as well. Terrorist groups are increasingly using new information technology (IT) and Internet to

- formulate plans,
- raise and launder funds,
- spread propaganda,
- communicate securely with the members (internal comm.) [10],
- share information and knowledge with similar groups (external comm.) [11],
- command and control [9],
- make research and development,
- recruit new members,
- generate international support,
- gather intelligence [12],
- make information warfare on behalf of the nations.

In addition to above, Internet offers;

- little or no regulation,
- potentially huge audiences,
- anonymity of communication,
- fast flow of information [10].

One of the striking examples that can be given on the use of websites by terrorists is PKK/KONGRA-GEL terrorist organization websites. There are 37 determined websites that are related with this organization. These websites generally include: the history of the organization, biographies of the influential people and its killed terrorists, and information on the political aims of the terrorist network. Content of these websites aims to create and enforce identity based separatism. One of the websites named “pajkonline.com” aims women who were mostly used in suicide bombings in the past. Another one is dedicated to cyber attacks and encourages the members to learn hacking techniques and provides information about the

vulnerabilities of computer operating systems [13]. This multidimensional example reveals the disrupting aspect of the issue.

IV. Cyber Terrorism Attempts and Further Expectations

Cyber attacks come in two forms; those that target data, those that target control systems [14]. Theft and corruption of data are the most common forms of Internet and computer attacks, and aim to sabotage services. On the other hand, attacks which focus on control systems are used to disable or manipulate physical infrastructure. For example, the provision of electrical networks, railroads, or water supplies could be infiltrated to have wide negative impacts on particular geographical areas. This can be done by using the Internet to send malicious programs or by penetrating security systems.

Weak spots of such an infrastructure were exploited in an incident in Australia in March 2000 where a disgruntled employee (who failed to secure full-time employment) used the Internet to release 1 million liters of raw sewage into the river and coastal waters in Queensland [14].

In 1998, a terrorist guerrilla organization flooded Sri Lankan embassies with 800 e-mails a day for a two-week period. The messages simply read “We are the Internet Black Tigers and we’re doing this to interrupt your communications.” Intelligence departments characterized it as the first known attack by terrorists against a country’s computer systems [4].

In July 1997, the leader of a Chinese hacker group claimed to have temporarily disabled a Chinese satellite and announced he was forming a new global “cracker” organization to protest and disrupt Western investment in China [10]. Internet saboteurs defaced the Home Page of, and stole e-mail from, India’s Bhabha Atomic Research Center in the summer of 1998. The three anonymous saboteurs claimed in an Internet interview to have been protesting recent Indian nuclear blasts [10].

Cyber terrorism may be used not only to inflict damage in itself, but in combination with conventional or nonconventional terrorism. Had Shoko Asahara and Aum Shinrikyo group been able to hack into the Tokyo power system and stop the subways, trapping passengers on the trains, the number of casualties caused by their 1995 Sarin gas attack might have been significantly larger [15].

Recently, Keith Lourdeau, deputy assistant director of the FBI’s Cyber Division, said the FBI’s assessment indicates that the cyber terrorist threat to the U.S. is “rapidly expanding,” and predicted that “terrorist groups will either develop or hire hackers, particularly for the purpose of complementing large physical attacks with cyber attacks” [16].

A survey of 600 IT security executives from critical infrastructure enterprises worldwide showed that more than half (54%) of them have already suffered large scale attacks or stealthy infiltrations from organized crime gangs, terrorists or nation-states. The average estimated cost of downtime associated with a major incident is $6.3 million per day [17].

The survey also introduced that the risk of cyber attack is rising. Despite a growing body of legislation and regulation, more than a third of IT executives (37%) said the vulnerability of their sector has increased over the previous 12 months and two-fifths expect a major security incident in their sector within the following year. Only 20% think their sector is safe from serious cyber attack over the following five years. 60% of those surveyed believe representatives of foreign governments have been involved in past infrastructure infiltrations. In terms of countries that posed the biggest threat to critical infrastructure security, the United States (36%) and China (33%) topped the list. More than half (55%) believe that the laws in their country are inadequate in deterring potential cyber attacks. Among the interviewees those based in Russia, Mexico and Brazil are the most sceptical. Among them 45% don't believe that the authorities are capable of preventing or deterring attacks.

Governance issues are at the centre of any discussion of security for critical infrastructure. Both the governments and private sector organizations need to gain cyber security capabilities. Although the security industry seems to stay one step ahead, governmental regulations has a vital role in defending critical infrastructures around the world. Moreover, a global cyber defense capability can only be obtained by means of international cooperation among the governments.

V. DEVELOPING INTERNATIONAL COOPERATION AGAINST CYBER TERRORISM

Starting with the basic "cooperation" thoughts, following part of the paper evaluates the international cooperation opinions from two aspects: Legislative cooperation and military cooperation.

A. The Legislative Cooperation against Cyber Terrorism

Up to today a number of both governmental and international steps have been taken. Governments are organizing themselves to confront the new threat. Some countries have established Computer Emergency Response Teams (CERTs) to handle incident response. USA and UK are the leading model nations for other countries that compose their cyber security policies. Many governments continue to struggle with the organization chart question, but some countries have been able to successfully form a national organization against cyber threats. For instance; in Brazil, the federal government established the Critical Infrastructure Protection Information Security Working Group, under its Department of Information and Communications Security in August 2009. This group works on information security and incident response plans. In Australia, a 2009 defense government report announced the establishment of a national Cyber Security Operations Centre, within the military's Defense Signals Directorate [17]. In Turkey, "The Scientific and Technological Research Council of Turkey" is tasked as a coordinator organization on cyber security. They have been able to form the national cyberspace security policy in 2009.

But the main problem is to establish a universal consensus on cyber threat. In recent years, a number of international communities have drawn the main frame and discussed the initial steps that need to be taken against cyber threat. However, global measures against cyber terrorism has not been addressed specifically yet.
European Commission Reports reveal the incoming threat as:
"The new information and communication technologies are having a revolutionary and fundamental impact on our economies and societies. In fact, the success of the information society is important for growth, competitiveness, and employment opportunities and has far-reaching economic, social, and legal implications. However, in the hands of persons acting in bad faith, malice, or grave negligence, information society technologies (ISTs) may become tools for activities that endanger or injure, the life, property, or dignity of individuals or even damage the public interest." [18]
"Despite the many and obvious benefits of the modern electronic communications development, it has also brought with it the worrying threat of intentional attacks against information systems and network platforms/infrastructures. As cyberspace gets more and more complex and its components more and more sophisticated, especially due to the fast development and evolution of (broadband) Internet-based platforms, new and unforeseen vulnerabilities may emerge." [19].
The European Union has therefore taken a number of steps to fight harmful and illegal content on the Internet, protect intellectual property and personal data, promote electronic commerce and tighten up the security of transactions. However, in spite of the EU initiatives, many observers believe that cybercrime requires an international response that should include countries that are havens for cybercriminals [20].
Council of Europe (CoE) Convention on Cybercrime released the first international declaration on crimes committed via the Internet and other computer networks. Four categories of criminal offenses are defined in the CoE Cybercrime Convention:

1) Offenses against the confidentiality, integrity, and availability of computer data and systems;
2) Computer-related offenses;
3) Content-related offenses;
4) Offenses related to infringements of copyright and related rights.

The purpose of this convention was "to make criminal investigations and proceedings concerning criminal offences related to computer systems and data more effective, and to enable the collection of electronic evidence of a criminal offence" [21]. In this regard, international legal measures play a critical role in countering cyber terrorism. Since the nature of the cyber terrorism issue is global then the response should be global as well. Globalization of crime demands globalized law enforcement [22].
The four criminal offenses defined above, confirm the ideas presented in the second part of this paper "The Term Cyber Terrorism". International common definition of the cyber terrorism should not only include destruction, degrading and denial with or through cyber- or IT-related means, but should also include the use of internet website contents for terrorist purposes.

In order to establish a global cooperation, a concerted strategy and policy should be constituted. There is a need to continuously watch, examine, observe and review terrorist organization websites. In order to maintain a common understanding and cooperation among international community, a consistent intelligence sharing and assembling process should be carried out. It is vital in our era, since terrorist organization websites and vulnerabilities of the ITs offer terrorists lots of opportunities for their activities.

Collecting intelligence is the starting point and the key part of building an international cooperation. Afterwards defensive and offensive (deterrence) collaborative actions should be set out. Counter information and cautions to the related public opinion and parties must be provided by the international organizations as defense strategies. The collection of electronic evidence whenever it relates to terrorism is crucial for the nations who desire to cooperate. An "Intelligence pool" should be created in order to collect and share the intelligence simultaneously among the nations. This intelligence pool should not only monitor and gather information from terrorist websites, but should also collect electronic evidence for the potential cyber attacks.

Knop, offers an "open source intelligent system" on this issue. Instead of a hierarchical organization, there should be a network, and knowledge should be pooled. There should be committee management, and a credit point system. Governments should be allowed to use the resource only to the extent that they contribute good quality information and analysis [1]. The collective open source idea is a well thought-out response to the challenge of organizing international cooperation regarding terrorist contents on the Internet.

"M.U.D." (Monitoring, Using, Disrupting) approach is also a well-organized applicable approach. According to MUD approach; "monitoring" forums, blogs and frequently updated terrorist websites gives information about terrorist organizations' motives, mindsets, audiences, operational plans and potential target population and potential targets for attack. In the "using" step, retrieved data can be achieved to identify the propagandists, members, connections between people and organizations. This approach also helps to identify the countries those support terrorists by means of funding and politics. Disrupting step can be applied by infecting the terrorist websites by viruses, worms and by destroying or changing the contents of the website [1].

MUD approach has many advantages. But disrupting step has the challenge of gathering all the participating nations on the same denominator. While one country wants to disrupt the content on a website, another country could still want to monitor and use that website in order to get more intelligence.

Monitoring and using steps could be organized to understand the radicalization process of the terrorist organizations. De-radicalization opportunities can be obtained after understanding the reasons of the radicalization. The target population of the terrorist organization could be reached and educated with a comprehensive human focused education program campaign on the web and media. This campaign should be developed and prosecuted for the various types of regional cultures.

According to Janczewski, building company defenses will not always be enough to reduce threats. Quite often a wider cooperation is required. This cooperation may be split into two streams. Stream one would group organizations using similar systems or facing similar threats. The best example would be the cooperation between Internet service providers (ISP). The handling of distributed denial of service attacks is much simpler if ISPs are working together on this issue. Stream two is to coordinate national and international law. Common sense dictates that if hacking would be made strictly forbidden in each and every country, then the number of hacking attacks would definitely drop across the globe [5].

Laws and conventions such as Council of Europe Convention on Cybercrime should be utilized to facilitate worldwide cooperation. Unless these conventions are expanded to include all the nations in the world, efforts will remain relatively inconclusive. But in order to respond to this kind of global threat, the key factor is to agree on a common definition of the threat. However implausible it may seem to reach a global consensus, there are many examples where such worldwide cooperations are already in effect. Air traffic control is an example of such global security arrangements.

Since the provisions of international agreements supersede the provisions for international cooperation, not only bilateral agreements but also multilateral agreements among nations must be signed. UN Security Council should also focus on cyber terrorism threat. Most of the permanent members of the Council are also the most vulnerable and targeted countries in the world. These countries also host most of the international cyber attacks. According to the charter of the UN, all members of the United Nations agree to accept and carry out the decisions of the Security Council. While other organs of the United Nations make recommendations to governments, the Council alone has the power to take decisions which member states are obligated under the charter to carry out. Therefore, worldwide cooperative law enforcement decisions against cyber terrorism could be taken under UN. Only under UN common definitions of "terrorism" and "cyber terrorism" could be generated and also "intelligence pool" against cyber terrorism could be formed. Since the nature of the cyber threat is global and spans throughout the world, the organization to respond this threat should be comprehensive and global.

A robust, international legal framework under UN that addresses cyber aggression is the most critical component of a comprehensive approach to deter cyber attack, much more critical than national offensive and defensive cyber capabilities. International law and norms are fundamental to deterrence because states "share an interest in adopting or codifying common standards for the conduct of international transactions…or in promoting or banning specific kinds of behavior by" states [23]. In this way, international law builds the framework that guides how and when states employ offensive and defensive cyber capabilities and forms the foundation of cyber deterrence. International law adds certainty to punitive actions and amplifies the costs of cyber attack by engendering a negative response from the international community, not just from the attacked state [24].

Unfortunately legislative measures are not adequate to fight against cyber terrorism. Military deterrence measures should be established in order to make

terrorists hesitate exploiting internet for their own destructive purposes. Proactive actions are required to disrupt the information on these websites and to locate and neutralize the attack's origin. In order to take offensive deterrence measures, NATO and other international organizations should establish deterrence strategies and keep agile and quick response teams always at their finger tips.

B. Military Cooperation against Cyber Terrorism (Cyber Deterrence and NATO)

The term "cyber deterrence" is the proactive measures that are taken to counter cyber terrorism activities. The mission of cyber deterrence is to prevent an enemy from conducting future attacks by changing their minds, by attacking their technology, or by more palpable means (such as confiscation, termination, incarceration, casualty or destruction) [25]. In response to a cyber attack, retaliation is possible, but is not limited to the cyber domain. For example, in the late 90's the Russian government declared that it could respond to a cyber attack with any of its strategic weapons, including nuclear [26].

NATO is the unique international military organization in the world that has cyber-defense and deterrence capability against cyber terrorism. The cyber terrorism against critical infrastructures and ITs is a growing threat for the member countries as well. Since the origin of an attack can be oversees, then it should be treated like an intercontinental ballistic missile attack. Moreover, the possibility of a large-scale cyber attack that comprises military force components is much more than the possibility of a ballistic missile attack. And also missile defense system will be able find adequate time to detect and engage the missile in seconds. But in the response to cyber attacks, there may not be sufficient amount of time to react. Therefore, NATO's next concept should cover cyber-attack defense shield, following the missile defense shield.

NATO's new Strategic Concept focuses on importance of terrorism and cyber-terrorism. According to Strategic Concept; cyber attacks are becoming more frequent, more organized and more costly in the damage that they inflict on government administrations, businesses, economies and potentially also transportation and supply networks and other critical infrastructure; they can reach a threshold that threatens national and Euro-Atlantic prosperity, security and stability. Foreign militaries and intelligence services, organized criminals, terrorist and/or extremist groups can each be the source of such attacks [27].

NATO aims to combine the cyber-deterrence abilities under centralized defense system. Strategic concept intends to develop further NATO's ability to prevent, detect, defend against and recover from cyber-attacks, by using the NATO planning process to enhance and coordinate national cyber-defense capabilities, bringing all NATO bodies under centralized cyber protection, and better integrating NATO cyber awareness, warning and response with member nations [27]. This is the first time that an international organization seriously declares its members are going to coordinate and cooperate their national cyber-defense capabilities.

Experts report on the new concept emphasizes that it is vital for NATO to respond to the rising danger of cyber attacks. Report reveals that NATO must accelerate efforts to respond to the danger of cyber attacks by protecting its own communications and command systems, helping allies to improve their ability to prevent and recover from attacks, and developing an array of cyber defense capabilities aimed at effective detection and deterrence [28].

According to report the most probable threats to NATO in the coming decade are unconventional. Three in particular stand out:

- An attack by ballistic missile,
- Strikes by international terrorist groups,
- Cyber assaults of varying degrees of severity.

Since the next significant attack in the near future might be expected from cyberspace, NATO has taken some steps to develop these capabilities through creation of a Cyber Defense Management Authority, a Cooperative Cyber Defense Centre of Excellence, and a Computer Incident Response Capability. Nonetheless, serious gaps still exist in NATO's cyber defense capabilities.

Already, cyber attacks against NATO systems occur frequently, but most often below the threshold of political concern. However, the risk of a large-scale attack on NATO's command and control systems or energy grids could readily warrant consultations under Article 4 and could possibly lead to collective defense measures under Article 5. Effective cyber defense requires the means to prevent, detect, respond to, and recover from attacks.

Utilizing NATO's Strategic Concept and the Experts Report on the Concept, a series of recommendations can be deducted.

- All NATO members should recognize that cyber attack is a growing threat to the security of the Alliance and its members.
- A major effort should be undertaken to increase the monitoring of NATO's critical network and to assess and furnish remedies to any vulnerabilities that are identified.
- The Centre of Excellence should help members improve their cyber defense programs through training.
- Allies should expand early warning capabilities in the form of a NATO-wide network of monitoring nodes and sensors.
- The Alliance should be prepared to send an expert team to any member country experiencing or threatened by a major cyber attack [28].
- Over time, NATO should plan to mount a fully adequate array of cyber defense capabilities, including passive and active elements.
- The Alliance should consider giving NATO military leaders certain pre-delegated authorities, based on agreed rules-of engagement, to respond in an emergency situation of a cyber attack [28].
- Member countries should establish their own cyber response teams, as well. Cyber defense and deterrence exercises that include different member and PfP nations should be held more frequently to train these quick response teams and share experiences on the issue. (e.g. Baltic Cyber Shield (BCS), as a highly useful international technical cyber defense exercise, was executed in May 2010. The exercise was organized

in collaboration with several organizations coordinated by Cooperative Cyber Defense Centre of Excellence (CCDCOE) and Swedish National Defense College (SNDC). An overall objective of the exercise was to gather lessons identified for the future cyber shield exercises planning [29].)

VI. RECOMMENDATIONS AND CONCLUSION

For all the reasons discussed above, it is an obligation to develop an international game plan in order to fight against cyber terrorism. Therefore, an 8-step global counter cyber-terrorism game plan is offered:

Step 1. Reaching to a common definition of terrorism and cyber terrorism is the starting point. Which activities on the internet (e.g. hacking, propaganda, attacking to infrastructures etc.) should be counted as cyber terrorism must be defined exactly. Speaking the same language or creating a common technical language could be a commencing point.

Step 2. Essential national and international legal measures have to be taken. International legal arrangements should be realized. Then national legislation has to be harmonized with the international legislation.

Step 3. Both bilateral and multilateral agreements on cyber security cooperation should be signed among nations.

Step 4. An intelligence pool should be created in order to collect and share the intelligence simultaneously among the nations. Collecting intelligence should include not only monitoring terrorist websites but also collecting electronic evidence for the potential incoming cyber attacks.

Step 5. Cyber defense expert teams should be created and charged internationally whenever a country encounters with a cyber attack. The number of quick response teams that countries own could be raised by the help of NATO's Computer Incident Response Capability and Cooperative Cyber Defense Centre of Excellence. An international counter cyber attack response training programme should be established.

Step 6. International counter-cyber attack exercises should be planned and executed in order to help the nations share their proficiency and experience.

Step 7. A well-organized international decision-making process that spans from detection to destruction (or disruption) of the cyber attack should be formed. Internationally authorized executives should respond to any attack concerning international security, based on agreed rules-of engagement.

Step 8. After-reaction analysis should be accomplished in order to identify and improve the weak points of the system. A feedback should be carried out for examining of the necessary innovations.

In consequence, cyber terrorism is a growing concern for the whole international community. The current regime of international laws, norms, and definitions not only insufficiently addresses cyber terrorism; it actually intensifies the dangers of the threat by creating a gray area or gap that can be exploited by cyber terrorists. Response to this global threat should be global as well. National efforts should be coordinated internationally to be successful against cyber terrorism. Countering

this threat requires not only legislative but also military cooperation including deterrence strategies. United Nations and NATO are two key international organizations. Due to UN's unique international character, and the powers vested in its founding charter, the organization can take action on a wide range of issues. Common definitions, international legal amendments and multilateral agreements might be considered and discussed under UN. And the steps concerning international military deterrence could be discussed under NATO and shaped under the guidance of Strategic Concept of NATO. It should be kept in mind that, international cooperation against global cyber terrorism threat is crucial and developing further proactive strategies for UN, NATO and other international organizations (e.g. European Union, Council of Europe, G-8, OECD) is essential.

REFERENCES

[1] K. Knop, "Institutionalization of a web-focused, multinational counter-terrorism campaign – Building a collective open source intelligent system, A discussion paper," Responses to Cyber Terrorism, Centre of Excellence Defence Against Terrorism, Ankara, Turkey (Ed.) IOS Press, 2008, pp.8-23.

[2] P. Wilkinson, Political Terrorism, London, 1974.

[3] M. Bozdemir, "What Is Terror and Terrorism?," School of Political Sciences Press and Publication College, 1981, v, vi. See also Wilkinson, P., (op. cit.), p. 17, and Crenshaw, M., 'The Concept of Revolutionary Terrorism', Journal of Conflict Resolution, September 1972, pp. 384.

[4] D. Denning, "Cyber terrorism. Testimony before the Special Oversight Panel on Terrorism," Committee on Armed Services U.S. House of Representatives, Georgetown University, May 2000. http://www.cs.georgetown.edu/~denning/infosec/cyberterror.html

[5] L. J. Janczewski and A. M. Colarik, Cyber Warfare And Cyber Terrorism, Information Science Reference, 2008.

[6] M. J. Warren, "Terrorism and the Internet," Cyber Warfare And Cyber Terrorism, Information Science Reference, 2008, pp.42-49.

[7] T. Oba, "Cyberterrorism seen as future threat," Computer Crime Research Centre Tech. Report, April 2004, http://www.crime-research.org/news/2003/04/Mess0103.html

[8] P.W. Brunst, "Use of the Internet by terrorists, A threat analysis," Responses to Cyber Terrorism, Centre of Excellence Defence Against Terrorism, Ankara, Turkey (Ed.) IOS Press, 2008, pp.34-60.

[9] Z. Sütalan, "Current and future trends in terrorism," COE-DAT Newsletter vol.3 issue.16 p.37-49, July-September 2010.

[10] K. Curran, K. Concannon and S. McKeever, "Cyber terrorism attacks cyber warfare and cyber terrorism," Information Science Reference, 2008, p.1-6

[11] M. Rogers, "The psychology of cyber-terrorism," Terrorists, Victims and Society, In A. Silke (ed.), Chichester:Wiley, 2003, pp.77-92.

[12] A. Silke, "The Internet & terrorist radicalisation: the psychological dimension," Terrorism and the Internet, IOS Press, H.-L.Dienel et al.(Eds.), 2010, p.27-39.

[13] E. Çelebi, "Analysis of pkk/kongra-gel websites to identify points of vulnerability," Responses to Cyber Terrorism, Centre of Excellence Defence Against Terrorism, Ankara, Turkey (Ed.) IOS Press, 2008, p.127-141.

[14] R. Lemos, "What are the real risks of cyber terrorism?," ZDNet, 26 August 2002.

[15] J. J. I. Noble, "Cyber terrorism hype," Jane's Intelligence Review, 1999.

[16] D. Verton, “CIA to publish cyberterror intelligence estimate,” ComputerWeekly.com. 2004. http://www.computerweekly.com/Articles/2004/02/25/200518/CIA-to-publish-cyberterror-intelligence-estimate.htm

[17] “In the Crossfire: Critical Infrastructure in the Age of Cyberwar”, A global report on the threats facing key industries, commissioned by McAfee and authored by the Center for Strategic and International Studies (CSIS), 2010.

[18] European Commission. (2001c). Communication on creating a safer information society by improving the security of information infrastructures and combating computer-related crime (eEurope 2002) [COM(2000) 890 final, 26.01.2001]. Brussels, Belgium: European Commission.

[19] European Commission. (2001b). Proposal for a council framework decision on combating terrorism [COM(2001) 521 final, 19.09.2001]. Brussels, Belgium: European Commission.

[20] S. M. Kierkegaard, “EU tackles cybercrime, cyber warfare and cyber terrorism,” Information Science Reference, 2008, p.431-438.

[21] International Working Group (2002). “Common position on data protection aspects in the draft convention on cyber-crime of the Council of Europe,” Retrieved December 15, 2004 from http://www.datenschutz- berlin.de/doc/int/iwgdpt/cy_en.htm

[22] S. Özeren, “Cyberterrorism and international cooperation: General overview of the available mechanisms to facilitate an overwhelming task,” Responses to Cyber Terrorism, Centre of Excellence Defence Against Terrorism, Ankara, Turkey (Ed.) IOS Press, 2008, p.70-88.

[23] C. W. Freeman, Jr., “Diplomatic Strategy and Tactics,” US Institute of Peace, 1997, p.84.

[24] S. W. Beidleman, “Defining and Deterring Cyber War,” Strategy Research Project, U.S. Army War College, 2009.

[25] T. J. Mowbray, “Solution architecture for cyber deterrence,” 2010, Retrieved January 11, 2011, from http://www.sans.org/reading_room/whitepapers/warfare/

[26] M. Libicki, “Cyberdeterrence and cyberwar”, 2009, Retrieved January 27, 2010, from http://www.rand.org/pubs/monographs/2009/RAND_MG877.pdf

[27] 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, 2010. Retrieved January 9, 2011 from http://www.nato.int/strategic-concept/index.html

[28] “NATO 2020: Assured security; dynamic engagement analysis and recommendations of the group of experts on a new strategic concept for NATO,” Experts Report on New Concept. 17 May 2010.

[29] Baltic Cyber Shield Cyber Defence Exercise 2010, After Action Report, 2010.Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

"Information Troops" – a Russian Cyber Command?

Keir Giles
Conflict Studies Research Centre
Oxford, UK
keir.giles@conflictstudies.org.uk

Abstract- **Appraisals of Russian military performance during the armed conflict with Georgia in August 2008 noted, among other deficiencies, poor performance in Information Warfare (IW). This led to calls in informed commentary for the creation of dedicated "Information Troops" within the Russian armed forces, whose duties would include what we would define as cyber operations. This stemmed from a perception in parts of the Russian Armed Forces that the "information war" against Georgia had been lost.**

No such entity has appeared in the Russian order of battle, but the public discussion and military comment is informative. Prospects for the appearance of "Information Troops" have been discounted both officially by the FSB and privately by Russian military officers. Arguments put forward against a unit of this kind include the unsuitability of servicemen for advanced cyber operations, and the ready availability and deniability of talented civilian volunteers. But at the same time Russia's EW troops are seeing their role and profile evolve in a manner which suggests they may be acquiring at least some IW capability.

The Russian approach to IW differs from our own, and there are specific perceived internet vulnerabilities which further affect the Russian approach to cyber operations, and prompt Russian pushes for treaty arrangements governing cyberspace.

This paper draws on unclassified open-source media and interviews with serving Russian military officers to consider the Russian military view of cyber operations as a subset of information war, and the prospects for creation of "information troops" (whether given this name or not) in the context of ongoing Russian military transformation. Informal links with volunteer and co-opted cyber forces are also considered.

Keywords: Russia; military; information warfare; doctrine;

I. "INFORMATION WAR" WITH GEORGIA

The brief war with Georgia in August 2008 prompted critical reviews of all aspects of Russia's performance and capabilities in armed conflict. For the most part, this criticism focussed on clear and unambiguous shortcomings in the conduct of kinetic military operations [1], giving impetus to the fundamental transformations which at the time of writing continue to grip the Russian Armed Forces. But one aspect of the conflict provoked far more nuanced and uncertain assessments; this was how Russia had acquitted herself in "information war" with Georgia.

Debates in the West over the nature of cyber conflict are followed with interest in Russia [2], but are not mirrored in the Russian public narrative. Considerations of whether cyberspace is the "fifth domain" for warfare, or simply is a common factor to the other four, do not feature in discussion visible in open sources, except in citations of Western thinking – in fact the word "cyber" is strikingly absent from home-grown Russian analysis, which tends to use the term only to describe US or Chinese activities [3]. Instead, the Russian view of "information war" (*informatsionnoye protivoborstvo*, *informatsionnaya bor'ba*, or increasingly commonly, *informatsionnaya voyna*) is a more holistic concept than its literal translation suggests, carrying cyber operations implicitly within it alongside disciplines such as electronic warfare (EW), psychological operations (PsyOps), strategic communications and Influence.

In other words, "Russia views cyber-capabilities as tools of information warfare, which combines intelligence, counterintelligence, maskirovka, disinformation, electronic warfare, debilitation of communications, degradation of navigation support, psychological pressure, and destruction of enemy computer capabilities [4]." At a time when the term has been written out of US information operations doctrine [5], "information war" is still alive and thriving in Russian security considerations [6].

Yet Russian analyses of the "information war" with Georgia failed to arrive at a consensus on whether that war was actually won or lost [7]. The rapid development of the portrayal of the conflict in Western media, and the mixed success of penetration of the Russian narrative of forced intervention in response to intolerable "genocide", were cited as evidence by both sides in the debate [8]. In addition, while cyber "campaigns" before and during combat operations in South Ossetia and Abkhazia were not alluded to as a component of Russian overall strategy, it was noted that their contribution to the Russian strategic aims was limited to the information domain – in other words, while elements of Georgian strategic communications were effectively suppressed, broader attacks (for instance on critical national infrastructure) were not in evidence [9][10]. Regardless of the final conclusion, the common perception among those writing in open sources about the information aspect of the conflict was that the performance of the Russian military in this area badly needed to improve.

II. VULNERABILITY

Specific historical factors relating to the Russian adoption of the internet and information and communications technology (ICT) give rise to a sense of vulnerability in this field, which serves only to exacerbate what British expert James Sherr called Russia's habitual "conspiratorial view about absolutely everything" [11].

For instance, failure to develop indigenous ICT and communications networks technology has led to extensive reliance on foreign-built systems – so a writer on information security can note that:

"The information security of the Customs Service of Russia is under the control of Slovenia and Germany (Iskratel), Russian power engineering enterprises and Gazprom have their security looked after by Germany (Siemens) and Sweden (Ericsson), Slovenia and Germany (Iskratel) and the USA (Avaya) make sure there are no accidents on the Russian railways, and now the USA and France (Alcatel) are to guarantee civic safety for us with the MVD... As for our defensive capabilities, it must be noted that the Russian Ministry of Defence does not have its own fixed communications network as in other countries but leases communications systems from Rostelekom. But the Rostelekom long-distance communications network is... wide open to the world."

In other words, "'Caution, The Enemy is Listening' is not just a warning you find on old telephones, but an objective reality. Their ears are in every home, every workplace, every military unit [12]."

The vast majority of Russian writing on cyber conflict is defensive in tone, and focussed on information security and information assurance. Although official Russia now views the activities of NATO and the USA with less alarm than during peaks of tension in the first decade of the 21st century, it remains the case that the stated aim of US information operations is "to influence, disrupt, corrupt, or usurp adversarial human and automated decision making while protecting our own [13]" – and despite careful avoidance by the USA of casting the Russian state in the role of an adversary in cyberspace, this language is mirrored in the Information Security Doctrine of the Russian Federation. This document, not updated since 2000, emphasises:

"the development by certain states of 'information warfare' concepts that entail the creation of ways of exerting a dangerous effect on other countries' information systems, of disrupting information and telecommunications systems and data storage systems, and of gaining unauthorised access to them [14]".

This defensive theme to public statements from Russia contrasts with US and British official discussion of cyber issues, where reference to defence against hostile cyber operations is balanced with references to considering *offensive* cyber

operations within a range of tools available to respond to attacks – as for example with British Minister of State for the Armed Forces Nick Harvey referring to "exploiting cyberspace to enhance our defence – including the capability to exploit the weaknesses of our opponents. Cyber capabilities may provide the kind of precise and tailored effects which a conventional attack cannot [15]." Mention of offensive cyber activity by the state is strikingly absent from Russian open sources.

Another distinctive aspect of consideration of information warfare in Russia is preoccupations in other spheres of information competition, such as the vulnerability of national culture to outside influences – perhaps understandable in a nation which, as Timothy L. Thomas puts it, is "armed mentally with the experience of losing an ideology at the end of the Cold War (described by some as 'World War III')" [17]. This is another facet of the holistic approach to information security in Russia, and this too is reflected in the Information Security Doctrine, which includes as threats:

"the devaluation of spiritual values, the propaganda of examples of mass culture which are based on the cult of violence, and on spiritual and moral values which run counter to the values accepted in Russian society [17]."

Thus in the Russian view, the information threat to be countered is a holistic one consisting of both hostile code and hostile content, and the threat is real and current – Russian doctrine emphasises the constant role of IW in peacetime as well as during hostilities.

The view of Dmitriy Rogozin, Russia's Permanent Representative to NATO, of recent NATO pronouncements on cyber defence is predictably colourful: "in spite of all of the Russian side's initiatives, questions having to do with cyber-security were not added to the list as a review of the Russia-NATO Council's common threats. This means that this topic was closed for Russia - they do not want to discuss it with us." Rogozin uses Stuxnet as an example to suggest that those countries whom he considers Russia's adversaries are "developing systems to suppress the cyber-nets of a potential enemy or to introduce to the software of civilian production (mobile telephones, for example) harmful programmes that can be activated at moment necessary for the West... It comes as no surprise, then, that the US has no strong motivation to sign any global treaties on not using cyber-weapons, especially not with Russia, which potentially could be the object of cyber-attacks [18]."

President of the Academy of Military Sciences Army Gen Makhmut Gareyev refers to "subversive information technologies of the West" being the root cause of disorder in the Middle East and North Africa in early 2011. "Internet networks were implanted in Egypt, Tunisia and Libya over a two-year period. It started with systematic training for communication checks, without direct calls for unlawful actions. At the right moment, a centralized order was issued across all networks for people to take to the streets." Gareyev pointed to a full-spectrum information threat

consisting of both code and content. “You know how this was done in Georgia, Ukraine and Kyrgyzstan and is now being done in the Middle East,” he continued, adding that the main instigator is the US National Security Agency, which “controls the radio-electronic situation and internet structures across the world... It has open and secret branches in many countries…Any attempt of relevant national structures to counteract these actions is immediately portrayed as violation of freedom of expression and human rights, causing various sanctions [19]."

Given their role and history, both Rogozin and Gareyev could reasonably be expected to take a conservative view on the immediate IW threat posed by the West to Russia. But the view that political change in North Africa came about as a result of a Western IW/cyber conspiracy, which could now be implemented against Russia, has also been expressed by President Medvedev. Speaking at a meeting of the National Anti-Terrorist Committee in February 2011, Medvedev 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 [20].”

III. “PLAYING CATCH-UP”

In keeping with a common perception that Russian security bodies moved from a very recent standing start in operations via the internet, the official history of the Institute for Cryptography, Communications and Information Technology (IKSI, originally training specialists for the FSB, SVR and other bodies, and now part of the FSB Academy) says that “test use of the Institute’s connection to the global Internet network” did not begin until February 1996 [21]. This was not long before, at parliamentary hearings entitled "Russia and the Internet: The Choice of a Future," FAPSI First Deputy Director General Vladimir Markomenko characterised the internet as a whole as a threat to Russian national security [22].

Certainly Russia’s first real exposure to “information war” involving public internet resources - countering Chechen information sources during the first Chechen war - was a sobering experience, and in the words of Paul Goble, “forced... Vladimir Putin to focus ever more closely on the role of the Internet in deciding the outcome of conflicts... Putin openly acknowledged that Moscow was playing catch-up on this battlefield: ‘We surrendered this terrain some time ago,’ he said, ‘but now we are entering the game again [23].’

After computer crime was defined for the first time in Russia’s 1997 Criminal Code, “combating crimes of this type became something entirely new for the law enforcement bodies. There were a lot of problems... the absence of practical experience or methods for investigating these crimes, or of a forensic system [24].”

Concerns about IW and cyber vulnerability continued to be expressed even before the armed conflict in Georgia. The then Deputy Chief of the General Staff, Lt-Gen Aleksandr Burutin, noted in January 2008 that "Russia should be ready for a global information war". "Leading states are now actively developing forms and methods of struggle in the information sector", since "the development of information technology transforms the idea of a state's military might and political potential, changes the traditional forms of power struggle... In the foreseeable future, the final aims of wars and armed conflicts will be achieved not so much by destroying the troops and forces of an adversary, as by suppressing its state and military command, navigation and communication systems, influencing other information facilities on which the stable government of a state depends [25]."

On the same day, Burutin said that information weapons which could be "used in an efficient manner in peacetime as well as during war pose great danger" for Russia. He voiced the Russian preoccupation with "the destruction of spiritual values, by targeting individual, group and mass conscience", noting that this was the area of activity of "a number of non-government organizations supported from abroad, to form a negative image of Russia [26]".

Despite the head of US Cyber Command, Gen Keith Alexander, describing Russia as a "near peer" to the USA in capability [27], this perception of vulnerability and sense that Russia may be lagging behind in development of official capacity for computer network operations (CNO) is reinforced by the dogged Russian emphasis on treaties or agreements to restrain the activities of states in cyberspace, and so-called arms control treaties for information weapons [28].

These efforts also involve the Ministry of Foreign Affairs of the Russian Federation (MFA) and Directorate K of the Ministry of Internal Affairs (MVD) [29]. Professor Igor Panarin of the MFA's Diplomatic Academy, the author of one of the standard works on Russian theory of information war [30], advocates "using the mechanisms of the UN and the mechanisms of Russian-American consultations to create new rules of the game, rules of information balance and rules for protecting our sovereign national information space [31]". It is argued that the 2009 agreement between Shanghai Cooperation Organisation (SCO) states on "cooperation in ensuring international information security", including provision for military cooperation, should be used as a template and extended [32]. Meanwhile, CSTO Secretary-General Nikolai Bordyuzha has said that his organisation too must "create a joint system to counter information threats", since: "a number of Western countries and international institutions, the first to step over the threshold of the information era, have stepped up the structural reconstruction of national and security systems on the basis of joining their information potentials into one to achieve political, economic, military and ideological dominance at the regional and global levels... Developing a common information space become particularly important. There is a need to create a joint potential for countering information threats, to secure information resources and communications of the CSTO bodies and the member states' national authorities [33]."

Taken together, this offensive on a broad front suggests strongly that Russia feels the need to complete its "catch-up" with foreign states, while further development by those states should ideally be limited by international binding agreements.

Some of the proposed treaty limitations make interesting reading when compared with anti-social behaviour in cyberspace which has emanated from the Russian Federation: Aleksandr Burutin backs "a mutually acceptable multilateral mechanism" which would bind states to "taking responsibility for what is happening in their information space" – a responsibility conspicuously absent in the case of Russia [34].

In treaty proposals as well as in doctrine, Russia conflates the threat from hostile bits with the threat from hostile content, which according to the Information Security Doctrine can "distort the perception of the political system, social order, domestic and foreign policy, important political and social processes in the state, spiritual, moral, and cultural values of citizens." It is for this reason, among others, that Russia is dissatisfied with initiatives proposed overseas - as uncompromisingly put by Khatuna Mshvidobadze of the Georgian Foundation for Strategic and International Studies (GFSIS), "Moscow refuses to sign the only promising agreement, the European Convention on Cybercrime, which has been open for signatures since 2001. The Kremlin does not want to cooperate with foreign law enforcement officials looking into something like the 2007 cyberattacks on Estonia, and it is surely does not want to risk exposure of its links" to cyber crime syndicates [35].

IV. "INFORMATION TROOPS"

When reviewing the military's performance in Georgia, deficiencies were noted in both the information-technical and information-psychological domains, the two main strands of information warfare in Russian thinking [36]. The answer, in the view of several informed critics, was the creation of "Information Troops" within the Russian Armed Forces, who would meet the military's need for full-spectrum information operations.

One of the most clearly developed arguments for an entity of this kind was put forward by Igor Panarin, referred to above. Panarin called for "Information Special Forces" who would "prepare for effective operations under potential crisis conditions [37]". These operations would cover all aspects of information operations, including CNO: as he noted elsewhere, "the objective is... certainly, to create centres which would envisage so-called hacker attacks on enemy territory [38]."

The holistic nature of the tasking for these new units, and the way in which the Venn diagram of the Russian information war concept includes much that we might categorise under entirely different headings, was illustrated by further

extensive and detailed descriptions of the desired new capability, which inter alia stated:

"The personnel of the Information Troops should be composed of diplomats, experts, journalists, writers, publicists, translators, operators, communications personnel, web designers, hackers, and others... To construct information countermeasures, it is necessary to develop a centre for the determination of critically important information entities of the enemy, including **how to eliminate them physically**, and how to conduct electronic warfare, psychological warfare, systemic counterpropaganda, and net operations to include hacker training [39]."

Persuasive press commentaries were followed in due course by Aleksandr Burutin noting at the National Information Security Forum that it was "essential to move from analysing the challenges and threats... to reacting to them and pre-empting them [40]". At the same time the Ministry of Defence acquired a new deputy minister specifically for information and telecommunications technologies, Dmitriy Chushkin [41].

V. COMPETITION

But when Col-Gen Anatoliy Nogovitsyn followed up by suggesting that the General Staff should be working on defence against information-technical attack, this military ambition was immediately criticised by the Federal Security Service (FSB): "It is a strange statement... Such issues are not under the purview of any one department and should be resolved within the framework of the country's Security Council" (a body saturated with serving and "former" FSB officers). "At the same time, the military cannot but know that we have already created information-protection mechanisms, and they are constantly being improved [42]."

This is indicative of the fact that this capability, which the military seems to feel it lacks, is already well-established in other of Russia's "power ministries" with permanent seats on the Security Council. The regulations on use of SORM, Russia's official monitoring system installed (and paid for) by ISPs, state that it is the FSB that accesses information on internet use on behalf of all other interested parties, or if they do not have sufficient technical means to do so, the MVD takes over [43]. The MVD has its "Directorate K" dealing with information crime in the broadest sense, and with a perceived ambiguous role in which kind of cyber crime it will prosecute and which it will leave in peace. Russia did at one point have a dedicated information security agency, the Federal Agency for Government Communications and Information (FAPSI) – described by one leading expert as "the unofficial Ministry of Information Warfare of the Russian Federation [44]". Although the life-span of FAPSI as an independent entity was relatively short, its components were not disbanded but absorbed into two other agencies – the Federal Protection Service (FSO) and the FSB [45].

While the FAPSI directorate dealing with government communications was transferred to the FSO [46], the FSB received the Main Directorate for Radio-Electronic Reconnaissance on Communications Networks (*Glavnoye upravlenye radioelektronnoy razvedki sredstv svyazi*, GURRSS). The influence of this body in directing policy today could be inferred from the fact that the former chief of FAPSI and of the GURRSS, Vladislav Sherstyuk, holds the information security portfolio on the Security Council and is also the head of the Department of Information Security at Moscow State University [47]. This department is particularly active in Russia's drives for international agreements on information and cyber conflict [48], referred to above. So a proposal for a new component of the Russian Armed Forces dealing with information warfare would have to contend with the fact that it would be launched onto a stage already crowded with other actors, who might be less than entirely willing to share space with a newcomer [48].

VI. THE REB TROOPS

Opinions on the prospects for "Information Troops" among senior Russian serving military officers interviewed for this paper vary widely. One dismissed the idea out of hand [50]; another expressed the view that although media chatter about "Information Troops" might be misguided, if a place were to be found for carrying out functions of the kind described within the Russian Armed Forces, it would be in the *Voyska radioelektronnoy bor'by, Voyska REB* – the Russian military's electronic warfare branch, to be translated here as REB Troops [51]. This was one of the few elements of the Russian forces whose performance did not suffer intense criticism after the armed conflict in Georgia (although as always, it is hard to distinguish Georgia's claims of effective enemy counter-measures from complaints that friendly communications systems simply didn't work in the first place) [52].

The emblem of the REB Troops, a spider astride a globe in the form of a latitude and longitude grid, is rich with temptation for those who would wish to interpret the symbol as meaning that operations using the internet are a key part of their role – even if there has been no explicit mention of formal expansion into cyber activities. (So much so, in fact, that the emblem has been appropriated by the self-styled "Cybernetic Police" for their website on information security and computer crime in Russia [53].)

Much of the upbeat material in open sources written about future plans for the REB Troops blends easily into the background noise of puff pieces about Russian military capability: they are not immune from the standard regular promises of new and improved equipment "which has no world equivalent". But at the same time, change does appear to be taking place there. Declaring a "REB Troops Day" alongside similar days for border guards, paratroopers etc. suggested a boost in status for the branch, even before a promise of reorganisation into an independent service arm in its own right [54], which if true would be a remarkable development. The REB Troops are currently part of the "Special Troops" (not to be

confused with "special-purpose troops" or Spetsnaz), i.e. troops with specialised functions which are not part of a force or service arm (*vid* or *rod voysk*).

At the same time the main role of the REB Troops has been re-defined as "winning and retaining superiority in command and control of combat actions" (a common phraseology in Russian definitions of information warfare), while "the effect of the actions of EW means are comparable with the use of modern high-precision weaponry". Furthermore, "in the near future fundamental changes in the development of EW means and materiel should allow it to develop into a specific main form of combat action, which in many ways will determine the course and outcome of armed conflict [55]." In short, although there is no direct evidence to support the suggestion that the REB Troops will be the locus of CNO for the Russian Armed Forces, the coloratura of official statements suggests that their role and prominence is developing in a new direction.

VII. DIY CYBER WAR

Another serving Russian interviewee was sceptical about the prospects for creation of effective "Information Troops", whatever their formal title might be, because of the difficulty of finding and retaining appropriate personnel within the military – in fact, he noted, the military was the wrong place for capabilities of this kind, since servicemen under orders could never compete in flexibility and creativity with civilian enthusiasts [56]. The tension between qualities desirable in servicemen and qualities desirable in "information warriors" may well be a universal problem - in the phrase of Brig-Gen Charles Shugg, Deputy Commander US 24th Air Force, it is hard to find people who are "military minded but still competent to be cyber professionals [57]". Just as with their counterparts overseas, the Russian REB Troops too report retention difficulties due to competition from civilian employers [58]. And in Russia, an additional constraint is imposed by reliance on conscription for a significant part of military manpower: it remains the case that those young males with an interest in, aptitude for, and access to ICT are the ones who are least likely to be conscripted, and therefore available for manning "Information Troops", because they are precisely the ones who have access to the vast wealth of online information explaining the best possible ways of avoiding the draft [59].

Yet according to one counter-argument, much of what the "Information Troops" would seek to achieve in terms of CNO need not be sited within the military at all. The cyber component of confrontation with Estonia in 2007 and Georgia in 2008, and the online assault against Kyrgyzstan in January 2009 [60], showed how little encouragement large sections of the Russian online community need to join in with furthering Russian state goals. A Russian survey of "actors in cyberspace" defines "Net NGOs" as "internet combatants who as a rule declare the absence of any link with State bodies but which as a rule are financed by them, or by other entities [61]". But with a light management touch ensuring that plausible (or even implausible) deniability is maintained, a nudge in the right direction is enough for campaigns rapidly to take on a viral nature. As suggested in one Russian report on

the cyber attacks on Georgia, “there is no need for the state machine in modern cyber warfare [62]”. When considering a loose network of highly technically capable individuals working towards a common goal, there is an obvious parallel with the Russian Business Network (RBN) cybercrime organisation [63].

With an overlap of tactics, techniques and procedures (TTPs) between cyber crime, cyber activism, and cyber aggression, from a Russian perspective the synergies are clear. As Alex Klimburg puts it, “the differences between these categories of cyber activity are often razor thin, or only in the eye of the beholder. From the perspective of a cyber warrior, cyber crime can offer the technical basis (software tools and logistic support) and cyber terrorism the social basis (personal networks and motivation) with which to execute attacks on the computer networks of enemy groups or nations.” Furthermore, “states have an interest in maintaining or tolerating proxy organisations that could be implicated in this type of activity and other forms of attack, such as distributed denial of service, which can be conducted by an average computer user with the right tools [64]”. Khatuna Mshvidobadze goes further and states that “the FSB's 16th Directorate is believed to control Russia's reserve force of hackers [65].” And in the words of the head of the Federation Council's Defence and Security Committee, Viktor Ozerov, briefing foreign military attaches on 18 March 2011, "there is still no special structure for countering cyber in the Armed Forces, but this does not mean that we are not dealing with these problems [66]."

The ready availability of cyber volunteers, or those who can be co-opted, is facilitated by the relatively low barriers to entry to would-be cyber miscreants in Russia. Some of the scripts and instructions distributed to aid those who wanted to attack Estonia in 2007 but didn’t know where to start may have been of an extraordinarily basic nature; but there is an impressively broad choice of Russian-language online resources available for the guidance and equipping of those who would like to develop their computer network attack (CNA) and penetration skills further. On the basis of the author’s entirely unscientific comparison, it appears a great deal easier (and cheaper) for a Russian speaker to find meaningful instructions, guidance and tools than for somebody seeking to make the same debut in English [67].

The concentrated power of deniable CNA operations from Russia is striking even when it is not directed abroad with hostile intent, as witness the fallout from the concentrated efforts to suppress the blogger Cyxymu in August 2009, when collateral damage meant large parts of Twitter, Facebook and LiveJournal were temporarily taken offline [68]. As David Hollis implies in the study cited earlier in this paper, when drawing lessons for future confrontations, in circumstances of this kind where the objectives of the perpetrators coincide precisely with the interests of the Russian state, is it important whether the aggressor party denies liability or not [69]? Much has changed since Moonlight Maze, when activities directed against US government computer systems reportedly ceased outside Russian office hours [70].

VIII. CONCLUSION

The narrative of "information war" is developing within Russia, but mostly under the influence of initiatives taken overseas. The approach to CNO by the USA and to a lesser extent by its allies is followed closely. The most recent senior comment on the subject at the time of writing came from influential long-term Duma deputy, and former Secretary of the Security Council and Deputy Minister of Defence, Andrey Kokoshin - a long-term proponent of the vital importance of information superiority for Russian security [71], with, intriguingly, a first qualification in radioelectronics from the then Bauman Higher Technical College [72].

Speaking at the launch of a report entitled "'Cyber Wars' and International Security" published in late January 2011 jointly by the Institute of International Security Issues of the Russian Academy of Sciences and the Faculty of World Politics of Moscow State University, Kokoshin said that "the development of issues of information warfare and 'cyber wars' must take place on an interdisciplinary level... the experience of many states shows that information warfare is not just a function of the Armed Forces: other state institutions including the secret services take part in it [73]". This makes an interesting counterpoint to the FSB statement cited earlier in this paper which appeared to be suggesting that it was not the business of the Armed Forces at all. The "'Cyber Wars' and International Security" report, according to the Russian Ministry of Defence newspaper *Krasnaya Zvezda*, "examines primarily US and Chinese policy in this area... The study examines issues such as operations in cyberspace as an integral part of information operations [74]." At the time of writing, the report itself appeared to be unavailable in open sources.

Meanwhile, Russian security concerns will continue to be prompted by the fact that "influencing the transfer and storage of data means that the physical destruction of your opponent's facilities is no longer required [75]" – potentially negating all the benefits of Russia's hard-won military reforms. Efforts will continue to be "directed at introducing international legal mechanisms that would make it possible to contain potential aggressors from uncontrolled and surreptitious use of cyberweapons against the Russian Federation and its geopolitical allies [76]."

So, Russian statements and initiatives on cyber operations have to be placed in this context of observing rapidly-developing capabilities overseas, and listening to public announcements in the USA and elsewhere of ever-greater potential and willingness to inflict damage on adversaries by means of cyber attack. At present, the urgent arguments for the creation of "Information Troops" within the Armed Forces have not yet given rise to any visible change in tasking or designation of military structures, and visions of Russia's potential organised cyber warriors range from the heroic and omnipotent [77] to the realms of surreal parody [78]; but there is no doubt that the preoccupation with a perceived lack of capacity to prosecute or defend against CNO within the military will continue to provoke calls for action.

ACKNOWLEDGEMENTS

Thanks are due to Timothy L Thomas of FMSO for exhaustive long-term research on Russian views on IW, extensively cited in this paper; and to Gordon Bennett for historical information on Russian security structures.

REFERENCES:

[1] "Understanding the Georgia Conflict, Two Years On", NATO Defense College, Rome, September 2010. Available at http://www.ndc.nato.int/research/series.php?icode=9

[2] BBC Monitoring: "Russia needs more cyber war specialists - prominent expert and Duma MP", Interfax-AVN, 1430 GMT 26 January 2011. Also Shavayev, A. G. & Lekarev S. V. "*Spetssluzhby i informatsionnoye prostranstvo*", *Razvedka i kontrrazvedka*, Moscow 2003, pp. 350-354, and Sharikov, P. A. "*Evolyutsiya gosudarstvennoy strategii v sfere informatsionnoy bezopasnosti*", *SShA – Kanada. Ekonomika, politika, kul'tura*, No. 12, December 2009, pp. 95-108.

[3] V. Shcherbakov. "*Prostranstvo virtual'noye, bor'ba real'naya*", *Voyenno-promyshlennyy kur'yer,* 13 October 2010; V. Sidorov. "*Kibervoyny: ot dozhdya k uraganu*", *Krasnaya zvezda*, 26 March 2008.

[4] K. Mshvidobadze, "The Battlefield On Your Laptop", Radio Free Europe/Radio Liberty 21 March 2011, available at http://www.rferl.org/articleprintview/2345202.html

[5] The 2006 revision of US Joint Publication 3-13, "Information Operations", "removes information warfare as a term from joint IO doctrine".

[6] T.L. Thomas. "Russian Views on Information-based Warfare", Foreign Military Studies Office (FMSO), July 1996; T. L. Thomas. "The Russian View Of Information War", FMSO, February 2000; T. L. Thomas. "Information Warfare in the Second (1999-Present) Chechen War: Motivator for Military Reform?", FMSO, 2002.

[7] P. Goble. "Defining Victory and Defeat: The Information War Between Russia and Georgia", in S. Cornell & F. Starr (eds) *The Guns of August 2008: Russia's War in Georgia*, New York 2009.

[8] M. Akhvlediani. "The fatal flaw: the media and the Russian invasion of Georgia", in P. B. Rich (ed.) *Crisis in the Caucasus: Russia, Georgia and the West*, London: Routledge 2010.

[9] A. Tsyganok. "*Informatsionnaya voyna protiv Rossii: kak eto bylo"*. Segodnya, 17 April 2009. Available at http://www.segodnia.ru/index.php?pgid=2&partid=13&newsid=8407.

[10] D. Hollis. "Cyberwar Case Study: Georgia 2008", *Small Wars Journal*, 6 January 2011. Available at http://smallwarsjournal.com/blog/journal/docs-temp/639-hollis.pdf

[11] "Russia: A New Confrontation?" House of Commons Defence Committee Tenth Report of Session 2008-09, 30 June 2009, available at http://www.publications.parliament.uk/pa/cm200809/cmselect/cmdfence/276/27602.htm

[12] A. I. Nogovitsyn. "*Za gran'yu informatsionnoy bezopasnosti*", *Zashchita i bezopasnost'*, No. 1, 2010. For an expression of similar concerns in a UK context, see A. Michael. *Cyber Probing: The Politicisation of Virtual Attack*. Shrivenham: Defence Academy of the United Kingdom, 2010.

[13] US Joint Publication 3-13.1, "Electronic Warfare".

[14] Available on the Security Council of the Russian Federation website at http://www.scrf.gov.ru/documents/6/5.html

[15] Public statement at Chatham House, 9 November 2010.

[16] T. L. Thomas. “Russian Information Warfare Theory”, op. cit.

[17] Information Security Doctrine. See also V. L. Sheynis. “*Natsional’naya bezopasnost’ Rossii. Ispytaniye na prochnost’*”, *POLIS. Politicheskiye issledovaniya*, No. 1, 2010.

[18] D. Rogozin. "The Price of the Issue", *Kommersant*, 16 February 2011

[19] Interfax news agency, 26 March 2011

[20] “*Dmitriy Medvedev provel vo Vladikavkaze zasedaniye Natsionalnogo antiterroristicheskogo komiteta*”, Russian presidential website, 22 February 2011, available at http://www.kremlin.ru/transcripts/10408

[21] *Kompant-dien, posvyashchennyy pyatidesyatiletyu IKSI*, Moscow: Institut Kriptografii, Svyazi i Informatiki, 1999; pp. 195-201.

[22] State Duma proceedings, 17 December 1996.

[23] P. Goble. “Russia: Analysis From Washington -- A Real Battle On The Virtual Front”, RFE/RL 11 October 1999. Available at http://www.rferl.org/content/article/1092360.html

[24] MVD website at http://www.mvd.ru/struct/10000220/10000288/

[25] ITAR-TASS news agency, 31 January 2008

[26] Interfax-AVN news agency, 31 January 2008

[27] “Cyber Threat to Pentagon is Global: China, Russia Near Peers of US”, 1 October 2010, http://www.geostrategy-direct.com/geostrategy-direct/secure/2010/10_06/ba.asp

[28] S. Gorman. “U.S. Backs Talks on Cyber Warfare”, *Wall Street Journal*, 4 June 2010. Available online: http://online.wsj.com/article/SB10001424052748703340904575284964215965730.html

[29] MVD website: http://www.mvd.ru/struct/10000220/10000221/10000740/

[30] I. Panarin. *Informatsionnaya voyna i diplomatiya,* Moscow: Gorodets 2004.

[31] BBC Monitoring: “Russian pundit interviewed on US information operations conference”, *Rossiya TV* 1950 GMT 27 April 2009

[32] S. M. Boyko, I. N. Dylevskiy, S. A. Komov, S. V. Korotkov. “*Voyenno-politicheskiye aspekty obespecheniya informatsionnoy bezopasnosti na prostranstve Shankhayskoy organizatsii sotrudnichestva*”, *Voyennaya mysl’*, No. 7, July 2010.

[33] “CSTO Needs Coordinated Information Policy – Bordyuzha”, Interfax 21 December 2010

[34] *Tsentr parlamentskikh kommunikatsiy*, 30 January 2009, available at http://www.parlcom.ru/index.php?p=MC83&id=27297

[35] K. Mshvidobadze, “The Battlefield On Your Laptop”, Radio Free Europe/Radio Liberty 21 March 2011, available at http://www.rferl.org/articleprintview/2345202.html

[36] T. L. Thomas. “Russian Information Warfare Theory: The Consequences of August 2008”, in S. Blank and R. Weitz. *The Russian Military Today and Tomorrow: Essays in Memory of Mary Fitzgerald*, Carlisle: US Army War College Strategic Studies Institute 2010.

[37] OSC: I. Panarin. "The Information Warfare System: the Mechanism for Foreign Propaganda Requires Renewal", *Voyenno-Promyshlennyy Kuryer* 15 October 2008.

[38] BBC Monitoring: “Russian TV highlights hacker attacks on Georgian sites”, *RenTV* 0930 GMT 11 November 2008.

[39] BBC Monitoring: “Russia is underestimating information resources and losing out to the West”, *Novyy Region*, 29 October 2008 (emphasis added). See also Tsyganok, op. cit.

[40] *Tsentr parlamentskikh kommunikatsiy*, 30 January 2009, available at http://www.parlcom.ru/index.php?p=MC83&id=27297

[41] Izvestiya, 27 February 2009

[42] D. Litovkin. “General Staff Prepares for Cyber War”, *Izvestiya*, 27 February 2009.

[43] For a detailed discussion of SORM, with legislative citations, see http://www.cyberpol.ru/sorm.shtml

[44] G. Bennett. *The Federal Agency of Government Communications & Information*, Conflict Studies Research Centre. Sandhurst: August 2000.

[45] G. Bennett. *FPS & FAPSI – RIP,* Conflict Studies Research Centre. Sandhurst: March 2003.

[46] Official history of the FSO, available at http://www.fso.gov.ru/histori/histori7.html

[47] Security Council of the Russian Federation website, http://www.scrf.gov.ru/persons/11.html

[48] See D. Talbot. “Russia's Cyber Security Plans”, 16 April 2010, available at http://www.technologyreview.com/blog/editors/25050/ for an interview with Sherstyuk discussing “cyber arms control” and the nature of cyber weapons.

[49] See also R. Heickerö, “Emerging Cyber Threats and Russian Views on Information Warfare and Information Operations”, Swedish Defence Research Agency (FOI), FOI-R-2970-SE, March 2010.

[50] Private interview, November 2010.

[51] Private interview, December 2010.

[52] R. Hamilton, “The bear came through the tunnel: an analysis of Georgian planning and operations in the Russo-Georgian War and implications for US policy”, in “Crisis”, op. cit.

[53] http://www.cyberpol.ru/

[54] “*Voyska radioelektronnoy bor'by stanut v armii RF samostoyatel'nymi*”, *Vesti.ru*, 15 April 2009, available at http://www.vesti.ru/doc.html?id=275300

[55] “*Sostoyaniye sil REB: interv'yu s nachal'nikom voysk REB VS RF O. Ivanovym*”, *Krasnaya Zvezda*, 15 April 2010.

[56] Private interview, January 2011.

[57] Speaking at “Cyber Warfare” conference, London 28 January 2011.

[58] “*Sostoyaniye sil REB: interv'yu s nachal'nikom voysk REB VS RF O. Ivanovym*”, *Krasnaya Zvezda*, 15 April 2010.

[59] For example the always-busy forum at http://www.antipriziv.ru/forum/ and many more.

[60] http://hostexploit.blogspot.com/2009/01/cyberwar-cyber-iron-curtain-now_28.html

[61] O. V. Kazarin, A. A. Salnikov, R. A. Sharyapov, V. V. Yashchenko. “*Novyye aktory i bezopasnost' v kiberprostranstve*”, *Vestnik Moskovskogo universiteta*: *Seriya 12, Politicheskiye nauki*, NN 2-3, 2010.

[62] BBC Monitoring: “Russian TV highlights hacker attacks on Georgian sites”, RenTV 0930 GMT 11 November 2008.

[63] For a well-constructed overview of RBN activities, see http://www.bizeul.org/files/RBN_study.pdf

[64] A. Klimburg. “Mobilising Cyber Power”, Survival: Global Politics and Strategy, vol. 53, no. 1, February-March 2011, pp. 41-60

[65] K. Mshvidobadze, “The Battlefield On Your Laptop”, Radio Free Europe/Radio Liberty 21 March 2011, available at http://www.rferl.org/articleprintview/2345202.html

[66] Interfax-AVN news agency, 18 March 2011

[67] For illustration, visit (with appropriate precautions) the fora at http://forum.inattack.ru/Barakholka-f11.html (for initial and advanced training) or http://forum.xakep.ru/forumid_307/tt.htm (for a bustling trade in tools).

[68] A. Michael. (2010) *Cyber Probing: The Politicisation of Virtual Attack*. Shrivenham: Defence Academy of the United Kingdom, p. 15.

[69] D. Hollis. "Cyberwar Case Study: Georgia 2008" in Small Wars Journal, 6 January 2011. Available at http://smallwarsjournal.com/blog/journal/docs-temp/639-hollis.pdf

[70] B. Drogin. “Russians Seem To Be Hacking Into Pentagon”, San Francisco Chronicle, 7 October 1999. Available at http://sfgate.com/cgi-bin/article.cgi?f=/c/a/1999/10/07/MN58558.DTL

[71] As cited in M. C. Fitzgerald. “Russian Views on Electronic and Information Warfare”, Hudson Institute, December 1996.

[72] Biography available at http://dic.academic.ru/dic.nsf/ruwiki/101812

[73] “Kokoshin: Kibervoyny ugrozhayut natsional’noy bezopasnosti Rossii”, One Russia party website, 26 January 2011. Available at http://er.ru/er/text.shtml?18/2254

[74] Ye. Podzorov. “*Ostorozhno, kibervoyny*”, *Krasnaya zvezda*, 29 January 2011. Available at http://www.redstar.ru/2011/01/29_01/1_02.html

[75] Prof. V. Lisovoy, speaking at Swedish Defence Research Agency, Stockholm 5 October 2010

[76] “Russian Federation Military Policy in the Area of International Information Security”, Moscow *Military Thought* 31 March 2007

[77] O. V. Kazarin, A. A. Salnikov, R. A. Sharyapov, V. V. Yashchenko. “*Novyye aktory i bezopasnost’ v kiberprostranstve*”, *Vestnik Moskovskogo universiteta*: *Seriya 12, Politicheskiye nauki*, NN 2-3, 2010.

[78] I. Koshkin. “*Zapiski o budushchey voyne – informatsionnyye voyska Rossii”. Voyenno-istoricheskiy forum,* 28 September 2009, available at http://vif2ne.ru/nvk/forum/archive/1758/1758322.htm

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

Is the Swedish Territorial Defence Ordinance applicable on the fourth arena?

Victoria Ekstedt
Legal adviser at the CNO Unit, Swedish Armed Forces
Enköping, Sweden
victoria.ekstedt@mil.se

Abstract- **Like other modern societies, Sweden is highly dependent on its digital infrastructure in order to run vital functions such as electricity, water purification, information and communications. Even though this infrastructure is characterized by transboundary features, it is clearly a part of the Swedish state. In peacetime, the Swedish armed forces are tasked to protect and defend the geographic territory of the state from violations, and the authority to do so is given by the Territorial Defence Ordinance. However, according to the analysis of this paper, the ordinance can not be applied on the digital parts of the society, by the military called "the fourth arena". Numerous difficulties rises with an application of the ordinance in its present wording and against this background, it is of interest to clarify the present legal situation and suggest a way forward in order to achieve adequate protection on the same premises as the other arenas. The interdependency between national and international law on this matter is pointed out and international law is used to interpret the national ordinance. The conclusion is that the Swedish politicians and legislators' needs to find legal support for the defence of the Swedish digital interests and infrastructure by seeking cooperation and legislative solutions in an international context and by doing so, hopefully the national legislation will follow. The legal challenges faced by Sweden are likely to be similar in several comparable countries, why this discussion should be of interest for other states and held in an international context.**

Keywords: digital territory, international law, violation, transboundary features

I. INTRODUCTION

The Swedish armed forces are tasked by the Swedish government to defend and protect the territorial integrity and national independence of Sweden in accordance with international [1] and national law [2-3].
The Swedish Territorial Defence Ordinance (the ordinance) [4] is the national legislation which authorizes the armed forces to defend the Swedish geographic territory in peacetime, at least regarding the three traditional military arenas (land, sea and air). Most states have national legislations with similar function to the ordinance since in a society founded on the rule of law, the authority for state officials to use force have to be expressly given and regulated by law in order to fulfill the requirements of legality and human rights . The ordinance is the national interpretation and implementation of the international law principles on state sovereignty and the right not to have its territory violated by other states [5]. It gives the armed forces the right and duty to defend the territory by using force if necessary, during *jus ad bellum*, the lawful resort to force in peacetime. In case of a war situation, *jus in bellum*, the ordinance ceases to apply and will be replaced by the Laws of Armed Conflict, LOAC [6]. It is undisputable that the armed forces are obliged to defend the Swedish digital infrastructure in case Sweden becomes involved in a war on the same legal prerequisites, *jus in bellum*, as the other arenas. Equally clear is that if a computer network attack is to be regarded as an "armed attack" in accordance with article 51 of the UN Charter, this would constitute a right to self defense in the same way as the other arenas. However, the threshold of an "armed attack" in a digital context is yet to be defined [7] and from international state praxis, we can conclude that no computer network attack have yet been categorized as an "armed attack" according to article 51. In addition, the difficulties to identify the perpetrators and/or the responsible state in conjunction with a strong political context have an aggravating effect on the development of praxis on this legal area.[1] This directs our main focus to peacetime incidents and conflicts, where the territorial ordinance is applicable, and not in the context of war or an armed attack. The question about the definition of a "digital territory" is however equally important both in a *jus ad bellum* as well as in a *jus in bellum* context.

The international legal framework from where the territorial ordinance origins were developed in an era when the digital revolution were yet to be born. The technical development and the information revolution bring with it an entirely new set of requirements which stretches beyond the existing legal labels and structure. To be able to deal with the digital dimension, the international legal framework have to be prepared to develop to avoid the risk of becoming outdated or end up being contra productive and thereby short-circuit itself. Swedish (and most other) military doctrine

1 For example, the attack on Estonia in 2007 were initially regarded as an article 51 attack by a statement by the Estonian Prime Minister Andrus Ansip, but it was later withdrawn by the Minister of Defence Jaak Aaviksoo after discussions with EU and NATO since NATO at that time did not consider cyber attacks as a clear, military action.

is more easily developed than laws, and embraces the fourth arena which is the same thing as we in daily language call the digital infrastructure of the society but in a military context. Computers, Internet, networks and other sorts of data form an integral part of our society and its importance as well as our dependence of it is steadily increasing [8]. The Swedish government and the Ministry of Defense are fully aware of this development, which becomes clear in the latest bill " A functional defense" [9]. The bill points out the increasing importance of the digital infrastructure and that the threats against it have risen. The risk of a full scale cyber war taking place is not regarded as the most plausible or imminent threat, instead focus is set on the risk for cyber incidents and conflicts of a lesser magnitude, including web based criminality. Today, there are numerous international examples of cyber incidents and conflicts, from large-scale computer network attacks such as those which took place in Estonia 2007 and in Georgia 2008, to the Stuxnet virus in 2010 [10] and computer network exploitations such as secret intrusions, espionage and numerous thefts of information which happened to Google in China [11] and the US Defense in 2010 [12] . NATO and EU circulated warnings to protect secret intelligence material due to cyber attacks originating from China, and it was concluded that the EU system was vulnerable due to the fact that security efforts were the responsibility for each member state [13]. The Swedish foreign minister Carl Bildt has emphasized in the Statement of Government Policy 2010 as well as in articles in national and international press [14] the need to acknowledge the importance of the threats against the digital infrastructure and the Internet. He underlines the importance to protect the freedom of speech and the need for enhancement of the security on the global digital arena, not necessarily by regulating the Internet but by finding ways of discouraging the perpetrators. Finally, the Swedish EU Commissioner Cecilia Malmström points out the importance for states to cooperate in order to effectively handle the transboundary nature of digital threats. She is presently working active to strengthen the capability within EU [15] as well as initiating a dialogue with NATO and the United States.

Against this background, it is of interest to clarify whether it is possible for the Swedish armed forces to apply the ordinance on the fourth arena in order to protect the digital parts of the society in peacetime, *jus ad bellum*. The conclusion is however highly doubtful and the Swedish politicians and legislators needs to find a solution to this dilemma.

II. THE TERRITORIAL DEFENCE ORDINANCE

According to the Swedish Territorial Access Ordinance [16], foreign state owned vessels and vehicles need to get an advance permission to enter Swedish territory except from emergency situations or for innocent passage.[2] The Territorial Defence Ordinance becomes applicable when a state owned subject violates these rules in any

2 The expression "innocent passage" means that vessels by sea may pass through the territory under certain circumstances, for example without any stops, passing at a certain distance from the Swedish territorial waters and such. This rule has its origin in the UNCLOS, United Nations Convention on the law of the Sea article 17-19.

way. An unauthorized presence on Swedish territory gives Swedish armed forces the right and duty to react, if necessary by using force.

III. A NATIONAL DIGITAL TERRITORY?

The geographic territorial borders of Sweden can be determined by treaties, conventions, maps and nautical charts. However, there is no definition on where the ”digital borders” are drawn. If we can not determine a territory, it will be equally difficult to say whether an infringement has taken place or not. Therefore, a prerequisite to apply the ordinance on the fourth arena is to have some sort of territory and borders defined. This in turn implies that the legislation can not be applied in its present wording due to the unique and significant circumstances the digital dimension carries with itself. It is to some extent possible to define a digital territory on the same conditions as for the geographic territory since the digital infrastructure has geographic connections such as wires and cables, servers, nods and the software used on these. The hard- and software which keep up the digital infrastructure of a state could in this sense constitute a Swedish digital territory where information can be created, stored and pass through. Unfortunately, this suggestion is associated with several problems. For example, state owned information is not always handled within the geographic borders of Sweden. A recent investigation [17] showed that almost twenty percent of the Swedish authorities and municipalities were at risk at having their e-mail bugged due to the fact that their spam filters were located in foreign countries by “cloud computing”. If information for example should be stored at a server in another country it is highly questionable whether it is defendable at all, at least by Swedish armed forces. But how important is it to talk about borders and territory in this context when the digital dimension care so little about geography? Well, since the legal framework can not simply abandon its present structure, it have to continue to be fastidious about borders and national territories and this in turn, means the need for a solution to the problem remains.

IV. PRINCIPLE PROBLEMS

The ordinance primarily deals with the presence of state subjects and not of individuals. A consequence with an application of this ordinance on a digital territory is that every subject that wants to access the Swedish digital territory would have to be identified and classified as a state subject or an individual in order to initiate the process of giving permission to enter or not. Of course, this is an impossible task to perform and it is against several principles. In addition, according to the international law principle on proportionality the actions taken by a state to protect and defend its territory are not unlimited; they have to be proportionate to the violation. It is doubtful whether such procedures would be acceptable according to this principle. Finally, the founding idea of the Internet that information should be free and accessible for everyone would be violated, and the extensive surveillance would be against several human right principles, such as freedom of speech and the right to privacy [18].

V. LEGAL PROBLEMS

The alternative to adjust the ordinance with the purpose of facilitate its applicability on the fourth arena is problematic due to the fact that this law origins from the international law regarding state sovereignty and the right for states not to have their territories violated by other states [19]. International law is not possible to change for a single state, but states can refine and develop it by interpretation and then make own political standpoints. There are numerous examples when states declare their opinion of the law in specific cases, for example Russia has declared that they equal computer network attacks with the use of weapons of mass destruction [20]. An adjustment of the ordinance to the special conditions of the fourth arena has to have a strong connection to the political official standpoint with regard to international law, and as long as this is not developed on the digital area, neither will the ordinance.

The ordinance is only applicable on violations in peacetime, *jus ad bellum*, which means it does not embrace violations which amount to an armed attack on Swedish territory, *jus in bellum* [21]. If the ordinance is to be used in its present wording, it is necessary to define the expression "violation" in a digital context. Comparing what would constitute a "violation" on the other military arenas, the ordinance would regard the mere unauthorized presence of a foreign state subject on Swedish territory as a violation. Using a strict interpretation in the digital context, a violation would be any foreign state actors' presence on the Swedish digital territory, for example entering a Swedish website without some sort of an advanced permission. Even if such analogy is well in line with the wording of the regulation, this is obviously not going to be practiced. However, if we choose to take a step away from the strict legal wording and interpret violation as something more intrusive than the mere unauthorized presence, "violation" would mean intrusions in digital areas not open for public access. Such intrusions with the purpose of gaining access to areas which contains data otherwise restricted in order to achieve information or to be able to navigate in otherwise closed networks, means we end up in a completely different set of laws. Still and only due to their digital features, these cases will not be handled by the ordinance or the Swedish armed forces, but by the Swedish police. These actions are primarily categorized as criminal acts and they are sorted legally under national criminal laws. However, at the same time, they are equally infringements of article 2.4 UN Charter if committed by a state. Digital intrusions could also be regarded as a use of force [22] in this sense, however not particularly serious due to the lack of the element of being "armed", i.e. no weapons are used, or because their effects are not severe enough [7]. In an EU context, these intrusions are also illegal actions according to the Council Framework Decision on attacks against information systems and the Council Framework Decision on combating terrorism [23].

Another complicating factor is that in case of a violation of a Swedish digital asset which is geographically situated outside the Swedish territory, neither the Swedish police, nor the Swedish armed forces have any right to protect or defend it. In these cases, the only alternative is to be reliant on the actions of the authorities of the

country where the asset is placed. Another feature, however equally difficult, is the situation of interdependency between states, i.e. when a state is dependent on the function of digital assets and infrastructures owned by another state. Sweden is for example operating one of the thirteen root servers in the world which provides the key element of the domain name infrastructure of the Internet. Operating the root server and keep it stable and secure is certainly of interest for other countries than Sweden, even though they have no jurisdiction to take actions in case needed [24].

VI. PRACTICAL PROBLEMS

Why not solve the problem by claiming a Swedish digital territory and then apply the ordinance on the fourth arena in its present wording? The argument to do so is simple – why should not general rules be applied on the digital parts of the society as well, especially with regard to its increasing importance. Unfortunately, this solution is too elementary due to the far reaching consequences it would bring and as we must take into consideration. For example, it would demand that the digital borders would have to be constantly monitored and that, taking into consideration the amount of traffic passing these borders every day would be an extensive on the verge of impossible, task to perform. Further on, the definition of the border would need constant and immediate updates since it would be defined by the existence and use of hard- and software within the geographical Swedish territory and that development is a constant and uncontrollable process. However, there is digital infrastructure which is indisputably placed within the geographic borders, and thereby could constitute at least some parts of a Swedish digital territory. Unfortunately, the crucial information needed in order to apply the ordinance is where the border of the digital territory is. This means that even if information on the digital geography is provided in part, it will be impossible to draw a complete borderline and defend it in the same way the other arenas are doing. If the borderline can not be drawn, there is also a risk that the society will have different levels of protection, or in worst case, not be defended at all, which is a problem. Due to this, an application of the ordinance in its present wording on a Swedish "digital territory" does not appear to be practically possible at all.

In summary, there are principal, legal and practical problems which speak against an application of the ordinance on the fourth arena. To make use of this ordinance on an arena where completely different standards and conditions prevails implies a number of undesirable consequences.

VII. ARE WE GOING TO PROTECT AND DEFEND SWEDEN INCLUDING ITS DIGITAL DIMENSION?

If the ordinance cannot be altered to fit or be applied onto the fourth arena, and no digital borders are possible to draw, does this mean the Swedish digital society shall not be protected and defended as the land, air and sea territories are against violations? Is this a fair consequence of having an inadequate regulation? The answer to this question is without hesitation that the whole society shall be protected and defended,

including its digital parts. This intention is clearly proclaimed by the politicians in the latest defence bill [25] issued by the government. The content of the bill is one step in a direction towards a possible solution by discussing the tasks for the Swedish armed forces as well as national security, sovereignty and independence in a global context. By pointing out threats as well as their solutions as something which no longer necessarily occurs or is to be protected and defended within the national borders, but in our immediate region and beyond, the government stretches the Swedish responsibility beyond the geographic territory. The ratification of the Treaty of Lisbon, the article 47.2 of the Treaty of the European Union (TEU) and the solidarity clause in the Treaty on the Functioning of the European Union (TFEU) article 222 strengthens this view and implicates that Sweden is prepared to comprehend and take responsibility for the security for at least the other member states of the EU. This view was also established by the "solidarity declaration" made by the Swedish parliament in June 2009 which states that Sweden will not be passive in case a catastrophe or an attack would occur in another EU member state or a Nordic country. This declaration covers both civilian as well as military crisis which is an advantage in a digital context, for example with regard to the problem of identifying the perpetrator. However, the prerequisites for such cooperation are still to be developed and since the specific purpose of the ordinance is to give the Swedish armed forces the authority to defend the Swedish territory, the applicability of the regulation is strictly national and will not provide adequate legal basis for such cooperation.

VIII. LEGISLATION AND/OR INTERNATIONAL COOPERATION – A WAY FORWARD?

A possible legal solution to the problem is to write a new and separate law, which defines what parts of the digital society to be considered as critical for its function and provide the Swedish armed forces with authority, strictly restricted to these parts, to defend it from violations in peacetime. There is already a law [26] which gives this authority to Swedish armed forces to guard specific objects and buildings as they consider need to be protected from unauthorized access by the public in order to prevent sabotage, terrorist actions and espionage, but it do not apply to the digital infrastructure. Creating a similar regulation for digital infrastructure which needs to be protected could be a way forward in order to ensure the function of the society and at the same time take international responsibility in order to strengthen the global digital security. Although, this is only a half-way solution since it will not be the same as defending the whole digital infrastructure of Sweden. It will only strengthen the protection of parts of the infrastructure in case of violations. In addition, the political history of Sweden makes it difficult for politicians to legislate against a strong public opinion that the armed forces should not be put at risk of using force against citizens on Swedish territory.[3] Due to this reason, the chances of having this type of legislation are highly uncertain. Another suggestion which already has been addressed is to put

3 The strict division of tasks between the armed forces and the police can be traced back to events that took place in Ådalen in 1931 where Swedish military fired at and killed 5 people at a demonstration.

the territorial questions aside and focus on efforts to strengthen international cooperation on these matters. One tool could be the development of the content of the solidarity clauses in the TEU and the TFEU. Additionally, with increasing interoperability and interdependency between countries, there should also be of interest to discuss if there are duties countries would owe to each other in protecting their digital infrastructure; however that question is too large to be addressed further in this paper.

IX. CONCLUSIONS

In summary, we can conclude that the ordinance can not be applied on the fourth arena due to principal, legal and practical problems. At the same time, there is a clear intention from the political leadership that the Swedish digital infrastructure shall be protected and defended. All modern societies are highly dependent on the function of their computers, networks and communications, which is a strong indicator that this ambition will not change. The unavoidable conclusion is that if the ordinance cannot provide a legal basis for the protection and defense of the Swedish digital infrastructure in peacetime, the legislators have to find a new solution, especially taking into account that the fourth arena operates under special circumstances, not always easily compared with the other parts of the society. Efforts have already been made by the politicians in creating security in a global context and this should continue to be a highly prioritized goal. If an international system of cooperation is established with the purpose to better protect the digital assets of states and a legal position has been worked out to ensure its function, less attention will be drawn to geographical territory issues, and surely national legislation will follow in order to meet the needs of the twenty first century instead of trying to adjust the features of the modern society to an outdated legal order.

REFERENCES

[1] Charter of the United Nations (1945). Available at http://www.un.org/en/documents/charter/index.shtml

[2] Swedish Constitution, Instrument of Government (1974:152) 10:9

[3] Ordinance with Instructions for the Armed Forces (2007:1266) §2

[4] Territorial Defence Ordinance, in full English translation the ”Ordinance (1982:756) concerning intervention by Swedish Armed Forces in the event of violations of Swedish territory in peacetime and in neutrality” in Swedish. “Förordning (1982:756) om Försvarsmaktens ingripande vid kränkningar av Sveriges territorium under fred och neutralitet, mm.”

[5] UN Charter (1945) article 2.1 on state sovereignty and 2.4 the principle about non-intervention between states, The UN General Assembly resolution 2625 (XXV) the “Declaration on Friendly Relations” (1970) and international customary law principles.

[6] The humanitarian laws (Geneva Conventions (1949) I-IV and Additional Protocol (1977) I and II), laws on Neutrality (Hague Conventions (1907) V and XII) and law on Occupation (Hague Convention (1907) IV)

[7] T. Wingfield, “When Is a Cyber Attack an" Armed Attack?": Legal Thresholds for Distinguishing Military Activities in Cyberspace,” *Potomac Institute for Policy Studies*, 2006.

[8] A thorough picture of our dependency of the Internet is given in the Council of Europe Secretariat report ”Internet governance and critical Internet resources” p.7-29 (2009)

[9] The bill ”A functional defence”, Ett användbart försvar” 08/09:140 p. 28

[10] N. Falliere, L.O. Murchu, and E. Chien, “W32. Stuxnet Dossier”, *Symantech Security Response*, vol. 3, 2010, pp. 1-64. Available at http://www.symantech.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf

[11] Available at http://googleblog.blogspot.com/2010/01new-approach-to-china.html

[12] Statement by the Director of the US National Intelligence, Dennis Blair at the annual threat assessment at the US Senate 2010.

[13] Evans, Michael, “Cyberwar declared as China hunts for the West´s intelligence secrets”, *the Times* March 8 (2010)

[14] Swedish Foreign Minister Carl Bildt “Tear down these walls against Internet freedom” in *Washington Post* January 25 2010

[15] “Commission to boost EU´s defence against cyber attacks” Directive IP/10/12/39. Available at: http://europa.eu/rapid/pressReleasesAction.do?reference=IP/10/1239

[16] Territorial Access Ordinance, Tillträdesförordningen (1992:118)

[17] Alert from the Swedish Fortifications Agency and Symantech “Myndighet slår larm om it-läckor” in *SvD* 2 february 2011

[18] European Convention of Human Rights (1950), articles 8 and 10. Available at: http://www.icrc.org/eng/war-and-law/treaties-customary-law/geneva-conventions/index.jsp

[19] UN Charter (1945) article 2.1 on state sovereignty and 2.4 the principle about non-intervention between states, The UN General Assembly resolution 2625 (XXV) the “Declaration on Friendly Relations” (1970) and international customary law principles.

[20] V.I. Tsymbal, “Kontseptsiya Informatsionnoi Voiny” (concept of information warfare) speech at RUS and US conference in Moscow 1995. For a complete overview see Swedish Defence Agency report “Emerging Cyber Threats and Russian Views on Information Warfare and Information Operations” March 2010 on the Russian view on these topics.

[21] Ordinance (1982:756) concerning intervention by Swedish Armed Forces in the event of violations of Swedish territory in peacetime and in neutrality 1§

[22] For a general definition of the expression “use of force”, see UN General Assembly resolution 3314 “Definition of Aggression”

[23] Council Framework Decision 2002/475/JHA of 13 June 2002 on combating terrorism and Council Framework Decision 2005/222/JHA of 24 February 2005 on attacks against information systems

[24] Council of Europe report “Internet Governance and critical Internet resources” p.9 (2009)

[25] The bill ”A functional defence” 08/09:140 pp. 8, 34-37

[26] Law on Protection, Skyddslagen (2010:305) (*authors´ translation*)

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011

Rationale and Blueprint for a Cyber Red Team Within NATO

An Essential Component of the Alliance's Cyber Forces

Luc Dandurand
Cyber Defence and Assured Information Sharing
NATO C3 Agency
The Hague, Netherlands
luc.dandurand@nc3a.nato.int

Abstract– **This paper provides the rationale and blueprint for a "cyber red team", a dedicated military capability whose objective is to improve the cyber defence of the Alliance through the controlled execution of cyber attacks. These cyber attacks would be specifically designed to achieve three goals. The first goal is to assess the effectiveness of the existing security measures in providing mission assurance, at both the technical and procedural levels. The second goal is to demonstrate the possible impact of these cyber attacks to senior management and key stakeholders. The third goal is to improve the cyber security staff's ability to detect and respond to cyber attacks by exposing them to realistic, unannounced attacks in their specific working environment. Details of the proposal cover governance, command and control, modus operandi, organizational structure, skills and experience required for team members as well as recommendations for personnel selection. It also identifies a number of controls that would address concerns related to its implementation.**

Keywords: NATO, cyber defence, cyber attack, cyber forces, red team, assessment, demonstration, training

This work was sponsored by NATO's Allied Command Transformation under the 2010 Cyber Defence Programme of Work. This document is a working paper that may not be cited as representing formally approved NC3A or NATO opinions, conclusions or recommendations, and represents only the views of the author.

I. INTRODUCTION

Given the immense complexity of large, modern communication and information systems (CIS), military organizations rely on risk management as the primary approach to achieve "adequate" security and protect the CIS from attack, each following the approach with a different degree of formality. Through risk management, military organizations identify security measures intended to bring the risk to a level thought acceptable without unduly limiting the usability of the protected CIS, an unfortunate side-effect of most security measures.

Many accelerating trends, such as convergence to IP networks, greater interconnectivity, and increased cyber threats, are resulting in greater uncertainty about the actual risks being taken. The cyber domain is simply changing too quickly for most military organizations to fully appreciate the impact of these changes on the security posture of their CIS. Best practices on risk management [1] are increasingly more difficult to follow under tighter constraints on time and budget, and the constant changes in operational requirements inherent to military missions. Finally, most of the security measures deployed in modern CIS require human involvement to function properly. Thus the effectiveness of security measures is dependent not only on the successful implementation of an underlying technical system, but also on the users' ability to operate it correctly and to follow specific processes.

The end result is that once a CIS is deployed, the senior decision makers responsible for its security and proper functioning, as well as those relying on it to execute their assigned mission, are sometimes left with a number of unanswered questions:

- Is the system sufficiently secure? Are some security measures unnecessary?
- Are the firewalls properly configured? Are the proper rules loaded into the intrusion detection system? Is the wireless network properly secured?
- Are the restrictions on the user workstations really necessary? Does it really help security to have new staff request access to each information resource independently? Does the single sign-on solution create a single point of failure?
- Will advanced, persistent threats be detected? Will potentially significant events be reported? Will all detected incidents be correctly analysed? Will staff know how to respond to an attack? Could security staff be overwhelmed by a cyber attack?
- What can an attacker do if he gains access to the CIS? How much information could be extracted before detection? If the attacker tried to modify operational information, would the users realize it before using the information? If the attacker destroys information, will it be possible to restore it from backups? How long will it take to restore each service?

Large military organizations require the capability to measure the actual effectiveness of the security measures deployed in operational CIS to provide mission assurance and reduce the uncertainty about the risks the CIS faces. As well, they require the capability to effectively demonstrate to senior decision makers the possible consequences of cyber attacks against specific military missions. Finally, given the significant degree of dependence of security measures on human processes, users and security staff require experience in how to respond to cyber attacks based on highly realistic scenarios conducted in their day-to-day environment, so that the cyber domain benefits from the same level of preparatory training as that being provided to the other domains of warfare.

Within NATO, the NATO Network Enabled Capability (NNEC) has begun shifting the traditional balance between security and ease of use and ease of access to information. "By improving collaboration in an open and dynamic information environment, NNEC enhances the efficiency and effectiveness of the Alliance." [2] For the International Security Assistance Force in Afghanistan, NATO's Afghanistan Mission Network is a significant step towards the NNEC. It reviews the balance between security risks and the benefits of an open and dynamic environment brought to overall mission effectiveness in a large coalition of military forces. More recently, the WikiLeaks incidents [3 and 4] have led some to question whether this new balance is the right one, and NATO's senior decision makers are trying to find the correct balance between sharing and protecting information given the realities of the modern cyber world.

This paper proposes the establishment of a "cyber red team" as a standing capability within NATO that would complement ongoing efforts aimed at addressing the above military requirements, as well as help NATO reap the full benefits of the NNEC with greater confidence that the actual risks being taken are in fact acceptable.

II. THE CYBER RED TEAM: AN ESSENTIAL MILITARY CAPABILITY

This paper considers a "cyber red team" (CRT) as a specific military cyber defence capability that provides a service to a requesting NATO organization. The capability and the service it provides to its "clients" are defined in some detail in order to provide a holistic and coherent view of how it could be properly managed, to dispel unfounded perceptions that it is a high-risk initiative, and to build confidence that, with the proposed control and accountability mechanisms, NATO can trust that the CRT will deliver the requested service in a proper fashion. Most if not all of the proposed implementation can be amended if necessary.

A. Mission of the Cyber Red Team

The mission of the proposed CRT is: "to assess the overall effectiveness of the security measures of an operational CIS in providing mission assurance through the controlled execution of no-notice, realistic cyber attacks, demonstrate their

mission impact to stakeholders and senior decision makers, and improve the cyber security staff's ability to detect and respond to these attacks".

As explained in Section I, the three activities identified in the mission statement (assess, demonstrate and improve) are the activities that could contribute the most to increasing the cyber security of NATO's CIS. The fact that the controlled execution of cyber attacks against operational CIS will generate factual evidence is the key value-added element for the assessment of security measures, at both the technical and procedural levels, and will enable credible demonstrations of the potential mission-level impact of these cyber attacks. As well, the fact that these controlled cyber attacks will be performed on operational CIS and without advising security staff in advance will allow for the best opportunity for improvement possible, short of an actual cyber attack.

It is important to note that the implementation of this capability will also provide insight into various aspects of cyber attacks, an element sometimes missing in the design and deployment of security measures. The rapid pace of change in the cyber domain requires defenders to remain abreast of the evolution of cyber attacks, and the CRT will provide critical information regarding the nature of cyber attacks to NATO and NATO Nations as a result of the execution of its mission.

1) Assessment of the Overall Effectiveness of Security Measures and Processes in Providing Mission Assurance

One of the three primary activities of the proposed CRT is to assess the actual effectiveness of security measures in an operational CIS and determine the extent to which they contribute to mission assurance. Such an assessment is performed only at the request of the head of a client organization, who will also define its scope and objectives. By their nature, these assessments will cover not only the technical and procedural aspects of security measures, but also how well they actually integrate together, a key aspect typically not verified by conventional security assessments. Given that human actions and processes play a fairly significant role in nearly all security measures, and given human nature, a realistic assessment of the overall integration of human processes and technical solutions can be achieved only if performed without notice, hence most staff members will not be advised of a CRT assessment. As well, the focus of the assessment is not the security measures themselves, but the impact of the cyber attacks on the mission given the effectiveness of the security measures.

A red team assessment is not a vulnerability assessment, nor is it penetration testing, as those terms are generally understood[1]. For most organizations, the former is generally undertaken in a collaborative fashion with the aim of listing all vulnerabilities in a network using automated tools. These tools typically show only the potential vulnerabilities of the systems assessed within the context and configuration of such systems, and do not indicate whether their exploitation is realistic given the system's configuration, the network's topology, its security

[1] The definitions of vulnerability assessment, penetration testing and red team assessment can vary from one organization to another, and thus there can be an overlap of the objectives and methodologies of these activities depending on which definition is used.

countermeasures, and the level of security monitoring. Nor do they provide any insight into what an attacker could do if he managed to exploit them. Penetration testing generally involves attempts to exploit possible vulnerabilities. It is a more comprehensive attempt at finding all vulnerabilities in a system, usually performed using highly specialized tools and custom scripts developed specifically for the targeted system. It is generally focused on a specific application or service, rather than an entire CIS, and is typically done collaboratively just prior to operational deployment. Thus the service provided by the proposed CRT is complementary to traditional vulnerability assessment and penetration testing activities.

Within NATO, the assessments to be performed by the proposed CRT would complement those already identified in the NATO Security Policy and Supporting Directives, with the additional advantage that the CRT activities would not be limited to assessing only vulnerabilities, but also the mission impact that can be achieved through their exploitation.

2) Demonstration of the Mission-Level Impact of Cyber Attacks

The second objective, demonstrating the impact of cyber attacks, will always be aimed at military operations or business processes at the mission level. The demonstration objectives will be determined by the head of the client organization, and will aim at showing the potential impact of specific cyber attacks to the organization's mission given the functionality provided by the operational CIS in support of that mission. Demonstration to stakeholders and senior decision makers is specifically mentioned as an objective because it is a key aspect typically not well addressed by most current security assessment activities, which are mostly focused on the technical functioning of CIS components. For example, a conventional assessment could determine that "it is conceivable that an attacker could exploit a newly discovered vulnerability in a cross-domain guard and gain access to an operational chat room and influence the command of military operations, but we think that we will detect that". Such a finding will never have as much value as "the CRT was able to force an infantry company to move from location A to location B during a training exercise by exploiting a vulnerability in a cross-domain guard, and did so without being detected". The fact that the activity is intended to remove uncertainty through actual demonstration of the possible impact allows senior decision makers to more objectively discharge their responsibility to balance security measures against competing CIS requirements such as ease of use, functionality, and the amount of investment and implementation time necessary for these security measures.

The only limitations on which effects can be demonstrated by the CRT are those brought by the necessity of maintaining control on these effects as well as on any second-degree effects resulting either directly from the CRT activities or from the reactions of staff not aware of such activities. Clearly, any improper manipulation of an operational military CIS can have serious consequences. While risks exist, they can be managed and maintained at an acceptable level at all times. This is the purpose of most of the controls described in Section IV.

3) Improving the Ability of Cyber Security Staff and Users in the Emerging Cyber Threat Environment

Proper functioning of most security measures depends to a large extent on their proper use by security staff and users. Users are given training on how to perform procedures that pose a risk to the security of CIS, such as transferring files using removable devices. Security-awareness programs educate users as to telltale signs of cyber attacks and advise them on how to handle them. Security staff are trained to detect and handle cyber attacks by operating specialized tools and following a number of processes that ensure detected attacks are correctly interpreted, stopped, and reported, and that necessary recovery actions are undertaken. Some of this training is given through dedicated courses, while some of it is achieved through cyber defence exercises. Courses are generally used to train security staff in the use of specialized tools, while exercises are generally used to train them in the execution of processes.

Although these training courses, awareness programmes and exercises are in place and definitely contribute to the overall security of NATO's CIS, they generally do not take place on the operational CIS, and they are not optimized for the day-to-day working environment of security staff. Finally, certain assumptions are often made regarding the outcome of business processes and/or whether security tools would have functioned properly, simply for efficient conduct of the training or exercise.

The controlled execution of cyber attacks against operational CIS will provide a clear opportunity for users and security staff to hone their skills with the tools they will use to handle real cyber attacks. Specific objectives can be defined to fill identified training gaps and to make sure that all staff are able to execute incident-handling processes correctly for the situations of concern. This is a key benefit provided by the CRT, since the quality of training obtained from courses and traditional exercises is very much limited by the level of reality of the training or exercise.

B. Cyber Red Team Tasks

The proposed CRT will fulfil its mission by undertaking each assignment within the context of a "task". A distinct task is created for each request for the CRT's services by a client organisation. The concept provides a logical framework that addresses key requirements for command and control, for defining the legal basis, and for information management. It also allows for the concurrent execution of multiple assignments by the CRT. To be effective, a CRT task needs to be executed over a period of between six and fifteen months. This is to ensure that the CRT has the opportunity to make a comprehensive assessment without substantial prior knowledge of the organization's CIS and internal processes and to properly demonstrate the impact of advanced, persistent threats.

1) Simulated Threats

For each task, the client will define in very general terms the threats the CRT must simulate. These can include a foreign intelligence service, a criminal organization,

an ideologically motivated hacker group, a malicious insider, and military forces with a computer network attack capability. Of course the reality of the simulated threats will be limited by the capabilities of the CRT, and the objective is simply to generally define the modus operandi of the CRT during the execution of the task and the types of activities it will perform.

2) Typical Activities

Each task will have a specific list of authorized activities, which could include:

- Gathering and taking advantage of public information from open sources
- Scanning and probing networks (wired and wireless) and telephone systems
- Performing social engineering
- Monitoring facilities, including "dumpster diving"
- Exploiting vulnerabilities and compromising client systems
- Exfiltrating information
- Conducting denial-of-service attacks against specific services or networks
- Modifying operational data
- Attempting physical access to facilities to gain access to CIS.

Clearly the above activities must be legally authorized, and they must be performed with sufficient controls to ensure that they will not cause unintended consequences.

III. GOVERNANCE, COMMAND AND CONTROL

Management of the capability has been divided into three levels: strategic, operational and tactical. The main reason for differentiating between the strategic and the operational levels is to create a clear delineation of responsibilities and thus contain liability in case an error or fault is committed by members of the CRT. The main reason for differentiating between the operational and the tactical levels is efficiency, as explained in Section III.D.

At the strategic level, a "Steering Committee" will direct the proposed CRT. The use of a Steering Committee addresses the requirement for having representation from both Strategic Commands and the civilian structure in order to support the CRT's NATO-wide remit and to ensure the CRT is independent of the CIS providers and their security staff. This is a key aspect of the capability, as a proper assessment cannot be provided by those responsible for the operation of a CIS or those responsible for the operation of its security measures.

At the operational level, control will be provided by a Task Control Team (TCT), defined specifically for each task. Since the CRT can execute multiple concurrent tasks, there can be several different TCT in existence at the same time. Finally, at the tactical level, "Attack Team Leaders" within the CRT will oversee actions taken by CRT staff members. An exact placement of the proposed CRT within the

overall NATO organizational structure has not yet been suggested. This is a secondary consideration given that regardless of its organizational location, it will report to a Steering Committee specifically established to support it.

A. Legal Framework

A full legal analysis of the implications of establishing and operating the proposed CRT is required, but is beyond the scope of this preliminary proposal. The two main issues identified at this point are the need to legitimize the CRT activities that could otherwise be construed as malicious or unauthorized use of computer systems, and the potential for invasion of privacy resulting from CRT activities. Since some NATO Nations are already performing red team activities in a similar manner as proposed herein, it is reasonable to expect that a suitable legal framework can be established.

B. Strategic Direction and Guidance

At the strategic level, a Steering Committee will be established to direct the CRT and guide its continuous evolution. It will have at least the following responsibilities:

- Maintaining mission and vision statements, defining key values and ethical behaviour for the CRT staff, setting the high-level objectives, priorities and milestones for the evolution of the capability over time
- Ensuring that a generic legal framework for the different types of activities to be performed by the CRT is established and maintained and that the CRT has the required set of processes and procedures in place for achieving its objectives without undue risk or liability
- Overseeing staffing of the CRT and ensuring it is properly resourced
- Securing continued funding for the capability
- Securing support required from external parties (facilities management, common services, etc.)
- Identifying possible clients and tasks, and promoting the capability in various forums
- Prioritizing, scheduling and authorizing tasks
- Defining the elements to be audited and the manner in which audits will be performed, and setting the performance standards against which the CRT will be assessed (see Section IV.E)
- Accepting the findings of audits performed on the CRT, and ensuring identified issues, if any, are resolved in a timely fashion.

C. *Operational Control*

Operational control of the proposed CRT will consist of authorizing and directing all activities performed on operational CIS during the execution of the CRT's mission. It is at this level that the responsibility for any mishap or unintended consequence lies. Operational control of the CRT will be performed by the TCTs. At minimum, a TCT will consist of the Head of the CRT and a representative from the client organization who has been delegated the required authority. Both of these individuals will have veto power on all decisions made by the TCT, and thus the CRT staff members will be able to perform only the activities that have been authorized by both.

Within the TCT, the key responsibility of the client representative will be to accept the risk posed by proposed CRT activities to the operational CIS on behalf of the head of the client organization. The key responsibility of the Head of the CRT will be to ensure that the CRT is capable of successfully executing the proposed activity, satisfying himself that the staff members are sufficiently trained, that the exploit tools have been properly tested, and that possible secondary effects have been properly identified to the client representative so that he has the appropriate information regarding the risk posed by the activity. If an unforeseen consequence occurs despite the CRT having full and accurate information from the client organization, it will be the responsibility of the Head of the CRT. If a consequence that was foreseen actually occurs and is not well received within the client organization, it will be the responsibility of the client representative.

D. *Tactical Control*

The activities authorized by the TCTs will be defined in a certain amount of detail. For example, "scan a range of IP addresses for services", "deploy to a site and identify wireless access points", "attempt compromise of the server at IP address A.B.C.D", or "perform a denial-of-service attack against IPs in the range A.B.x.y". The amount of detail provided will be at the discretion of the TCTs. In most cases however, there will remain latitude in the specific execution of the activity, if only because a TCT simply will not be able to oversee every detail of a task. This "tactical control" will be the responsibility of the Attack Team Leader (see Section V.A). The Attack Team Leader will control and oversee the staff within his team and ensure that activities are executed in accordance with the direction provided and all applicable procedures. He will also take part in most activities, and will be responsible for constant oversight of the operational activities. Finally, he will also be responsible for ensuring the targeted CIS can be restored to its original state at the end of the task.

IV. MANAGING THE OVERALL RISK

The activities to be performed by the proposed CRT pose certain risks, including:

- Actions on a target system could cause unintended effects, such as rebooting it, affecting the functioning of services, or causing the loss of data

- Actions on a target network could cause unforeseen, collateral consequences, such as affecting dependent systems that were not to be targeted, consuming a substantial amount of bandwidth, destroying data, or triggering alerts
- Staff from the client organization could detect the red team activities and react to them in a problematic manner
- A red team staff member could act maliciously during a task.

To properly address these significant risks, specific controls have been built into the proposed CRT. The following sections provide insight into the most important of these controls.

A. *Trusted Agents*

In order to ensure staff members within the client organization do not react in a problematic fashion to detected CRT activities, and to ensure that unforeseen consequences are detected in a timely fashion, it will be necessary to place "trusted agents" at key positions within the client organization. Trusted agents will be identified when a task is initiated and will be given 24/7 contact information for various members of the CRT. They will be provided with sufficient information to allow them to immediately identify activities that could potentially originate from the CRT. When such activities come to their attention, they will contact the CRT who will in turn advise them of how to properly handle the situation.

Trusted agents will be selected from key positions within the client organization along the incident-handling process from sensor to decision maker to operator. A sufficient number of agents will be required to ensure that the CRT will have enough "eyes and ears" at the client organization to detect in a timely fashion any potential problem that may result from its activities.

In some cases, the CRT may choose to inform trusted agents of a specific action in advance in order to adequately control the risk it poses. For example, if the CRT is tasked to modify information in an operational command and control application to verify whether its users or security staff would detect attacks against the integrity of the information, a sufficient number of trusted agents in the immediate vicinity of the targeted users would be kept in direct and constant contact during the activity to ensure none of the users reacts to the modified information in a problematic fashion.

B. *Authorized Activities*

The CRT will be able to perform only those activities that have been specifically authorized prior to execution. There are two levels of prior authorization. The first will occur in the development of the task plan, in which the types of activities to be performed are identified in general terms, reviewed by legal counsel and authorized by the head of the client organization. These become "sanctioned" activities.

The second level will occur in the actual execution of the task, when the TCT authorizes that one of the sanctioned activities be performed in a specific way. The TCT may or may not make use of all sanctioned activities, but under no circumstance will it be able to authorize an activity that has not been sanctioned in the task plan. This second level of authorization will ensure that the TCT agrees that the specific activities are aligned with the objectives of the task and do not pose an unacceptable risk.

These authorizations would be of no use if they were not backed up by a code of conduct that each team member must agree to follow in order to join the team, the means to detect attempts by a team member to perform unauthorized actions, and administrative regulations that would enable NATO to take the required actions against such an individual. These are also addressed within the proposed CRT.

C. *Sanctioned Targets List*

Another key control mechanism is the Sanctioned Targets List (STL). The STL is a controlled document, signed by the head of the client organization, that will identify the systems that can be targeted by the CRT during a task. The STL can also specifically list the types of activities that can be performed against each target, thus further limiting the scope of authorized activities. It will be the responsibility of the client organization to ensure that it has authority over all of the systems listed in the STL.

D. *Two-Person Rule*

To address the risk that a CRT staff member could act maliciously and abuse the access to systems and information achieved by the team during a task, a two-person rule will be put in place. The two-person rule requires that all actions taken on an operational CIS be performed by two staff members working together. The proposed CRT will make use of a dedicated facility for the conduct of operations specifically to accommodate this rule (see Section V.B). Any team member observing another member working alone in the dedicated facility will have the obligation to challenge that person and report any suspicious activity to the TCT. The two-person rule implies that at least two staff members would need to collude to perform malicious activities, thus significantly reducing this risk.

E. *Comprehensive Auditing*

Another mechanism to address the risk that a CRT staff member could act maliciously is the comprehensive use of automated auditing and a comprehensive review of all aspects of the CRT's activities by a Task Audit Team (TAT) appointed by the Steering Committee and the head of the client organization. All activities performed on operational CIS will be audited in a number of ways:

- All authentications to sensitive information stores and systems will be logged

- All custom-developed systems to support CRT operations will have significant auditing in place that monitors the access of CRT staff to sensitive information
- Keystroke loggers will be installed on key systems
- Full packet capture systems will record all traffic in and out of the CRT
- All input and output on shells on key workstations will be copied to log files
- Screenshots of key workstations will be taken at random intervals and recorded
- All logs will be centralized and available for the TAT's review.

Clearly, with this amount of comprehensive audit logging, there is a strong probability that malicious activity, if suspected, will be detected by reviewing the log material. While the depth of the review of the logged material will be left to the discretion of the TAT, the mere fact that so many of the staff members' activities will be logged will act as a significant deterrent to malicious activity.

F. Test Procedures

The CRT will be obliged to test all exploit tools and software it will use against or on the client's CIS. The testing must provide reasonable assurance that the software will not cause unintended consequences. It will be the responsibility of the Head of the CRT to ensure that minimum testing standards are clearly defined, and it will be the responsibility of the Attack Team Leader to ensure that all software used during a task has been properly tested according to these standards.

G. Management of Client Information

During the execution of a task, the CRT will obtain and generate a large amount of information about the client organization, some of it potentially highly sensitive. The CRT will ensure that this information is properly secured according to the applicable NATO and national policies. In addition, the CRT will limit the information it retains after a task to a "task record" used for the purposes of programme management, and general findings that can be re-used for cyber awareness programs within NATO, as specifically authorized by the client organization. All other information will be destroyed at the end of a task, and the destruction will be audited. Finally, in addition to keeping as little information as possible, the CRT will follow specific procedures to ensure that proper care is taken when handling personal information in order to protect the privacy of individuals.

V. ORGANIZATIONAL STRUCTURE AND SUPPORTING FACILITIES

A. Organizational Structure

Figure 1 shows the organizational structure of a red team suitable for the proposed mission. The two main components of such a team are the Attack Group and the Support Group, which are further described below.

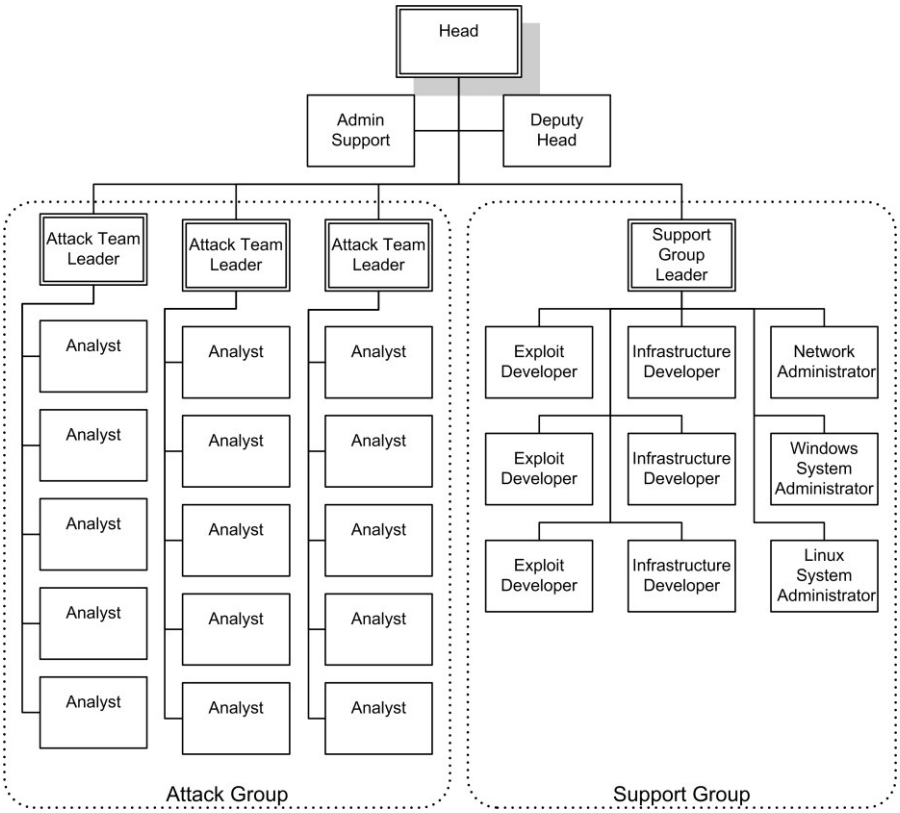

Figure 1. Organizational structure of a red team

The optimal size of a red team depends on a number of factors, such as:

- The size and complexity of the various CIS to be assessed
- The scope and depth of the assessment, the impacts to be demonstrated and staff improvement to be provided
- The frequency of the assessments
- The threat to be simulated and the level of desired reality.

While the size can vary according to these factors, the structure should remain the same as it is built around the different types of work and the skills and knowledge required in each area, as detailed below. There is however a bare minimum size below which it will not be possible to have the breadth and depth of knowledge required to provide realistic assessments, demonstrations and improvement. This bare minimum is 12 full-time members for the type of red team described herein. Even at 12 members, there remains a risk of mission failure, particularly if the staff selection process does not deliver the highest-calibre cyber security experts. It is important to note that mission failure in this case implies that the CRT would inaccurately assess the effectiveness of the security measures of an operational CIS, eventually leading senior decision makers and stakeholders to falsely believe that these measures are adequate.

The team size presented in Figure 1 represents the ideal size for the proposed CRT. The size of the team is based on the requirement to have a representative from each NATO Nation in addition to the three NATO civilians holding the positions of Head, Deputy Head and Admin Support. It is proposed that the national representative positions within the CRT be staffed through Voluntary National Contributions so that each Nation retains an agreed level of control and insight into the team's activities through a national chain of command.

1) Attack Group

The Attack Group is composed of three Attack Teams each consisting of an Attack Team Leader and five analysts. They are responsible for performing the required attacks against the targeted CIS and seeking ways of achieving the objectives of the task while staying within the defined boundaries. Analysts require a mixture of skills and experience:

- Expert knowledge of cyber security
- Strong knowledge of system and network management
- Strong experience in the Unix and Windows environments
- Strong knowledge of common Internet protocols
- Strong knowledge of wireless networks
- Military experience or at least a strong understanding of how military forces employ their CIS
- Ability to think "outside the box" and persevere.

2) Support Group

The Support Group is responsible for the development of the various tools needed by the Attack Group and the maintenance of the CRT systems. In addition to system and network administrators, the Support Group has two types of developers: Exploit and Infrastructure. Exploit Developers require:

- Expert knowledge of cyber security

- Software-development experience and reverse-engineering experience for both Windows and Unix systems
- Ability to search for vulnerabilities in software
- Ability to exploit buffer overflow and heap vulnerabilities
- Ability to code in assembly for different architectures.

Searching for vulnerability and developing exploit code is perhaps the most difficult work in cyber security. It requires a special "mindset" and extraordinary concentration, and these positions will likely be the most difficult to staff.
Given the requirement for custom software to provide a suitable exploitation infrastructure (e.g. "backdoors" and "covert channels"), efficient information management tools, and the required automated auditing and control mechanisms, the CRT requires an internal team of Infrastructure Developers who possess:

- Expert knowledge of cyber security
- Software-development experience for both Windows and Unix systems in several languages
- Experience in advanced version control and release management
- Experience in systems and network programming
- Experience in database development
- Experience in web development.

B. Physical Facilities

The conduct of red team tasks must be seen by all team members as a special activity. In addition to specialized systems, it requires concentration, focus and oversight. The proposed CRT would therefore perform its task from a purpose-built "Operations Room". This Operations Room would be designed specifically to address the human factors associated with the controlled execution of cyber attacks on operational CIS, accommodate the size of the team, and allow for proper demonstrations to senior-level decision makers.

C. Personnel Selection

A cyber red team is an elite team. To be successful, its members need to possess knowledge in a large number of highly technical areas. In addition, they need perseverance and an ability to think "outside the box". While all of these are the typical traits of a "hacker", the stereotypical hacker will also have the undesirable traits of disrespect for rules and desire for fame. Disrespect for rules and desire for fame are the most critical threats to the success of a professional red team. The leader of the CRT will play a critical role in establishing and maintaining the correct "mindset" among the staff, founded on meticulousness, rigour and discipline, in order to deliver a highly professional military capability. The CRT

also requires a strong team spirit, as it must achieve effects beyond those within the reach of individuals.

While common recruiting tools and methods can be used to screen applicants in terms of their education, experience and knowledge in technical areas, it is very difficult to evaluate whether candidates also possess the right attitude and "mindset". To create the best possible team, the following recommendations are made:

- The team's leader should have a few years' experience in managing a red team.
- The team should be built progressively so that the leader can instil the right values, attitude and team spirit in the members without being overwhelmed with new recruits.
- The selection process should allow for a multi-day competitive evaluation of a handful of potential candidates previously screened for their suitability in terms of education, experience and knowledge. The evaluation should be performed through a realistic simulation of a red team task and assess the candidates' abilities while under pressure for extended periods of time.

VI. CONCLUSION

The proposed cyber red team would provide a significant contribution to the improvement of NATO's cyber defence capability by identifying potential gaps and shortfalls in both technical solutions and incident-handling processes, demonstrating the mission-level impact of cyber attacks, and improving to the highest degree possible the skills and ability of security staff. Its implementation represents a significant, dedicated effort by NATO to perform an unbiased, highly realistic self-assessment of the effectiveness of security measures in providing mission assurance, and helps identify the most cost-effective way of improving NATO's cyber defence. Finally, it would also provide NATO Nations with insight into how cyber attacks can be successfully executed, and the mission-level impact these attacks can have against modern CIS.

REFERENCES

[1] G. Stoneburner, A. Goguen and A. Feringa, *Risk Management Guide for Information Technology Systems*, Special Publication SP 800-30, National Institute of Standards and Technology, 2002.

[2] "Allied Command Transformation NATO Network Enabled Capability Information Portal on TRANSNET", Internet: transnet.act.nato.int/WISE/Informatio, May 18, 2010 [Mar. 17, 2011].

[3] "The War Logs. An archive of classified military documents offers views of the wars in Iraq and Afghanistan", Internet: www.nytimes.com/interactive/world/war-logs.html [Mar. 17, 2011].

[4] Nick Davies and David Leigh, "Afghanistan war logs: Massive leak of secret files exposes truth of occupation", Internet: www.guardian.co.uk/world/2010/jul/25/ afghanistan-war-logs-military-leaks, Jul. 25, 2010 [Mar. 17, 2011].

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

Towards Establishment of Cyberspace Deterrence Strategy

Dmitri Alperovitch
McAfee Inc.,
dmitri_alperovitch@mcafee.com

Abstract— **The question of whether strategic deterrence in cyberspace is achievable given the challenges of detection, attribution and credible retaliation is a topic of contention among military and civilian defense strategists. This paper examines the traditional strategic deterrence theory and its application to deterrence in cyberspace (the newly defined 5th battlespace domain, following land, air, sea and space domains), which is being used increasingly by nation-states and their proxies to achieve information dominance and to gain tactical and strategic economic and military advantage. It presents a taxonomy of cyberattacks that identifies which types of threats in the confidentiality, integrity, availability cybersecurity model triad present the greatest risk to nation-state economic and military security, including their political and social facets. The argument is presented that attacks on confidentiality cannot be subject to deterrence in the current international legal framework and that the focus of strategy needs to be applied to integrity and availability attacks. A potential cyberdeterrence strategy is put forth that can enhance national security against devastating cyberattacks through a credible declaratory retaliation capability that establishes red lines which may trigger a counter-strike against all identifiable responsible parties. The author believes such strategy can credibly influence nation-state threat actors who themselves exhibit serious vulnerabilities to cyber attacks from launching a devastating cyber first strike.**

Keywords: cyberdeterrence, cyberspace, strategy, first-strike, counter-strike, confidentiality, integrity, availability

I. Introduction

Deterrence is a psychological 'game of chicken' that attempts to influence the cognitive state of the potential adversary actor and prevent them from embarking on course of action that they may wish to take. US Department of Defense defines

deterrence as the 'prevention from action by fear of the consequences. Deterrence is a state of mind brought about by the existence of a credible threat of unacceptable counteraction[1].' This is accomplished through either a directed or latent coercion that convinces the opponent that the costs of action are extremely prohibitive. Much of the strategic deterrence theory, which had been developed in the aftermath of World War II and enhanced over the duration of the Cold War, had focused on nuclear deterrence, achieved through the overt or opaque threat of nuclear force retaliation. This paper examines the traditional strategic deterrence theory and it's application to deterrence in cyberspace, which is the newly defined 5th battlespace domain, following land, air, sea and space domains and is being used increasingly by nation-states and their proxies to achieve information dominance and gain tactical and strategic economic and military advantage[2]. The paper does not attempt to address tactical cyberdeterrence for attacks that do not cause strategic damage to national or economic security, although it is possible to foresee a 'death by a thousand cuts' scenario, where numerous smaller cyberattacks can amount to a strategic threat.

II. DETERRENCE THEORY

There are two essential components to any viable deterrence strategy. In order for deterrence to be effective and credible, one must convince the potential adversary that you possess both the **capability** and the **will** to either retaliate or initiate a first preemptive strike to thwart an eminent attack[3].

The retaliation capability comprises of the ability to timely detect a threat (before the counter-strike assets or the C2 (command and control) to launch them are destroyed in the attack), rapid C2 decision-making and execution to launch a retaliatory or preemptive strike and ability to inflict prohibitively costly damage on the aggressor through such a strike. This capability can be demonstrable and overt, as with the unconcealed nuclear triad forces of the United States and Soviet Union during the Cold War, or unannounced and opaque, as with the widely assumed but unacknowledged nuclear arsenal of the state of Israel.

The will to retaliate, on the other hand, is a more nebulous psychological concept that is comprised of reputation (such as past willingness to use nuclear arms for the United States or historic proclivity to not put significant value on human life for the Soviet Union) and declaratory first-strike or second-strike policy. In the case of Israel's nuclear deterrence strategy, while there is no declaratory policy of nuclear retaliation due to its policy of ambiguity, there is a declaratory 'Shoah-proof' or 'Never Again' security doctrine which is designed to influence an adversary's thinking on the willingness of the country's leaders to use its undeclared nuclear arsenal as the last resort scenario[4].

III. TAXONOMY OF CYBERATTACKS

The rules of deterrence do not change once we move from a nuclear domain to a cyberspace one. The goal remains to influence the opponent's evaluation of one's capability and will to retaliate in response to an imminent cyber threat in order to dissuade them from ever launching the attack.

In order to evaluate what capability must exist for effective cyberdeterrence, we must first examine the types of cyberattacks that are in the realm of possible and analyze which ones have potential to be deterred.

For over two decades, the CIA (Confidentiality, Integrity, Availability) triad has been used in industry and academia to model the fundamental principles of information security[5]. It states that the goals of an information security system are to provide Confidentiality, Integrity and Availability of information. Conversely, cyberattacks can be classified using the same model as they exploit one or more of these attributes as they attempt to either steal information, or attack the Confidentiality part of the triad, modify information, or attack its Integrity, or prevent access to information, attacking its Availability.

IV. CONFIDENTIALITY ATTACKS

Confidentiality attacks (also referred by the US Department of Defense to as CNE, or Computer Network Exploitation) are nothing more than traditional espionage achieved through high tech means. Most of the sophisticated cyber attacks that are seen launched by either nation states or criminal groups fall into this category. History has shown us that espionage, known as the second oldest profession, has been around for nearly as long as there has been a human civilization and is an act that, while considered to be a sign of unfriendly relations, has become an internationally accepted norm that typically does not trigger more than a diplomatic retaliatory response[6]. Even during the darkest days of the Cold War and while faced with pervasive Soviet espionage activity in its most sensitive national security area – the 'Manhattan Project' initiated to design, create and test a nuclear weapon, the United States had not considered any type of retaliation beyond criminal prosecution of the spies and occasional expulsion of diplomats. It is widely acknowledged that even friendly nations spy on each other and when such activity is detected, it rarely has any effect, other than perhaps a fleeting chill placed on diplomatic relations. The international norms of just war theory dictate that retaliation must be proportional to the harm suffered from the attack[7]. Thus, it is unimaginable to envision a non-pariah state on the international scene responding with an overwhelming destructive attack to a case of cyber-espionage activity, no matter how damaging the loss of information had been to vital national security interests. And without the threat of massive response, the key pillar of the deterrence strategy is removed, preventing effective deterrence of confidentiality attacks in cyberspace.

V. INTEGRITY ATTACKS

Attacks on integrity are much more insidious as they are designed to achieve a tactical or strategic advantage over an adversary by sabotaging the operation of their critical civilian or military information systems. The sabotage can involve manipulation of data inside information systems that can degrade or distort the situational awareness capability of the adversary by spreading misinformation inside their intelligence systems with either a tactical objective to obscure specific activities that may be under surveillance or to achieve a strategic surprise in preparation for an attack. It can also involve subversion of physical devices and processes that are guided or operated by information systems, such as manipulation of weapons guidance systems to cause them to fire off-target. Targets can also include civilian critical infrastructure resources, such as electric grid, stock market and other financial databases, water filtration plans and others. The Stuxnet worm, discovered in June 2010, is believed to be the first publicly known nation-state sponsored integrity cyberattack, which is speculated to target the Iranian uranium enrichment program for subtle and long-term sabotage with the goal of destroying Iran's centrifuges by covertly making them spin at faster frequencies than they had been designed to do[8]. It is quite clear that these types of integrity cyberattacks pose a severe danger to advanced nation-states, whose economies, critical infrastructure and military systems are dependent on information systems, as they can be used to remotely wreak havoc on financial, energy, food, water, and transportation infrastructure sectors, as well as degrade abilities of advanced militaries to collect and analyze reliable battlefield intelligence and even execute kinetic operations. Deterrence of these types of attacks must be a priority for any effective cyberdeterrence policies.

VI. AVAILABILITY ATTACKS

Availability attacks are those that attempt to bring information systems offline in order to shutdown or destroy critical physical or virtual processes or prevent access to information. Long-lasting attacks can cause devastating damage to the economy, such as those that cause prolonged electricity or communication network blackouts. Short duration attacks that are surgically targeted at intelligence collection and analysis capability can blind a nation's ability to see an immediate strategic conventional or broader cyber threat by denying defenders access to vital situational awareness data or intelligence resources. Thus, just as with integrity attacks, availability threats can, under certain circumstances, present a serious national security danger and must be deterrable in a broader deterrence strategy.

VII. DETERRENCE IN CYBERSPACE

To achieve deterrence against integrity and availability cyberattacks, according to the deterrence theory, requires that a nation-state first build a credible threat detection capability that will protect its ability to counter-strike. While advanced cyber threats can use advanced obfuscation and polymorphic techniques to avoid detection for a prolonged period of time (as Stuxnet had done, avoiding all public detection for at least 12 months since its earliest proven sighting in the wild in 2009), the chances of them avoiding discovery permanently are quite low as their sabotage or other destructive activities will likely bring attention to themselves sooner or later.

Attribution is another problem that presents a challenge to the detection capability. It is very difficult and, often, impossible to accurately and quickly attribute a cyberattack, once it is discovered, to a specific adversary through technical means alone. The anonymity of the Internet easily allows an attacker to lay a false trail and hide behind a myriad of intermediately hop-points or proxy actors. The use of traditional non-cyber intelligence resources, including HUMINT and SIGINT, can help with that goal but they also cannot provide a reasonable level of assurance required for credible deterrence that the attacker will be identified. Despite this challenge, it is conceivable that deterrence will be effective even without accurate and timely attribution. For one, the required level of attribution required for a counter-strike is directly proportional to the degree of criticism a nation-state is prepared to endure in the international and domestic courts of public opinion and is greatly influenced by the destructiveness of the original threat and a cui bono analysis of the attack objectives. The threshold is not proof beyond reasonable doubt in the court of law but sufficient mix of suspicion and evidence to justify the retaliatory strike to the plurality of domestic and international audiences. For instance, strategic context of international relations at the time at which a cyber attack may take place can offer strong clues as to its origins[9].

One of the other major challenges one faces in applying lessons learned from traditional strategic deterrence theories to cyberspace is timely detection of the attack. Due to the nature of cyber weapons, cyber offensive capability can be developed, tested and pre-deployed offline without any credible means for detecting it. Unlike physical weapons, there is no missile silo, mobile launch facility or submarines to observe and monitor for early-warning detection. The attacks themselves may propagate literally with a speed of light when deployed through fiber optic networks, but even on slower copper networks, the speed of fully automated cyber ordinance will still be faster than what a human can reasonably evaluate and respond to. Deployment of defensive measures with automated retaliation capability, on the other hand, presents too high of a risk for misfire or targeting of an innocent party due to attribution challenges.

Accurately and timely determining the intent of the attack, once it has been discovered, is yet another problem. If international norms that dictate

proportionality of response prevent retaliation to attacks on confidentiality, or espionage, determining whether the goal of an intrusion is to attack confidentiality, integrity or availability of information system becomes nearly just as critical as detecting the attack itself. The process of intrusion classification can be very challenging for a defender, at least in its initial stages, as the tactics, techniques and procedures (TTPs) may be identical for all types of attacks. For instance, attacks on confidentiality, integrity and availability may all begin with an intrusion exploiting a vulnerability in an externally connected information system, followed by malware deployment, which provides the adversary with full remote access capabilities to the internal network. Until that remote access capability is leveraged for an integrity or availability attack, it may be impossible to know through technical means the true purpose of the intrusion. This uncertainty further adds to the complexity of establishing an early-warning detection system that can provide sufficient time for the appropriate Command and Control (C2) decision makers to evaluate the information and launch a counter-strike.

Advanced defensive tactics, technologies and highly trained personnel will contribute to the shrinking of the detection and classification gap. Separation of defensive and offensive resources, such as storage of offensive cyberweapons in offline locations which are less vulnerable to virtual targeting and distributing the retaliatory information systems and networks across wide virtual and physical space will help to build credible resilience to the counter-strike force. This can reduce the reliance on rapid detection and classification of inbound attack by providing the means for the decision makers to retaliate even after suffering a devastating first strike, minimizing the chance that the adversary can count on taking out all of the counter-strike assets in a single attack.

Second, is the need to preserve a rapid C2 decision-making and execution of a counter-strike option when facing a devastating cyber attack. This can be accomplished by preserving the resiliency and integrity of command chain communications by instituting or preserving offline communications channels that are less likely to be impacted by cyber attacks, such as dedicated traditional secure POTS (plain old telephone service) lines and encrypted radio and satellite communications that are physically separated from virtual networks which can carry attack codes.

Third, the counter-strike itself must be capable of instituting devastating damage on the attacker's own virtual and physical infrastructure to make the first-strike prohibitively expensive. Limited public demonstrations of cyber offensive capabilities can serve a useful purpose in alerting potential opponents to what they may face should they decide to attack. However, this part of the deterrence equation presents the biggest challenge to developed nation-states with advanced cyber defensive and offensive capabilities but who face developing nation-state adversaries with dangerous offensive cyber weapons but are themselves not reliant on cyberspace for their national economic or military interests. It is hard to cause

prohibitively devastating damage on your opponent through cyber means alone if his vital infrastructure is completely disconnected from the network. This problem presents a serious conundrum to policy makers, who face the unappealing choice of rising up the escalatory ladder and retaliating with conventional or perhaps even nuclear weapons in response to a cyber-only attack, in the process risking violations of international norms of proportional response, or absorbing the attack without a response and looking weak to their enemies, friends and populations alike. Yet, while this is a significant unresolved policy problem today, it is reasonable to expect that its consequences will lessen with time, as more and more developing countries rapidly increase their reliance on cyberspace in order to reap the economic, efficiency and force-multiplier benefits it affords.

Lastly, political leaders must demonstrate the credible will to issue a cyber counter-strike in response to a highly damaging integrity or availability attack of national security consequence. In today's world of complete ambiguity with regards to cyber-offense that can create much uncertainty in the minds of potential opponents, this can be best accomplished with a declaratory policy that defines, even if in opaque terms that provide sufficient room for decision makers to maneuver, the red lines that will trigger a counter-strike or even a preemptive first-strike in response to a credible and imminent threat.

VIII. CONCLUSION

This paper has argued that an effective cyberdeterrence strategy must focus on consequential integrity and availability cyber attacks and deter them through a declaratory policy that establishes red lines which may trigger a counter-strike against all identifiable responsible parties. It is also advantageous to provide public limited demonstration of offensive capabilities, spend resources on increasing threat detection and resiliency of both defensive and offensive networks, increase off-line redundancies for C2 communications and enhance HUMINT and SIGINT intelligence collection and analysis efforts to focus on cyber threat actors, their capabilities and intentions. This strategy can credibly influence nation-state threat actors who themselves exhibit serious vulnerabilities to cyber attacks from launching a first strike.

IX. FUTURE EXPLORATION

This paper did not explore several areas that need to be covered by a comprehensive cyberspace strategic deterrence doctrine and that should be explored in future works. These areas include the challenge of deterring non-state actors, such as terrorist and criminal groups, a problem that has not been adequately solved in neither the physical nor the virtual world. Another aspect that should be examined in the future is whether tactical as opposed to strategic cyberdeterrence is achievable against attacks that don't rise to the level of strategic impact.

ACKNOWLEDGMENT

The author would like to acknowledge Greg Conti, Adam Meyers, Jose Nazario, Jason Shepherd, and Jeff Stambolsky for their valuable and thoughtful contributions to the ideas expressed in this paper.

REFERENCES

[1] Joint Publication (JP) 1-02, *Department of Defense Dictionary of Military and Associated Terms*, 8 November 2010 (as amended through 31 January 2011), p107, http://www.dtic.mil/doctrine/new_pubs/jp1_02.pdf

[2] R. Rozoff, *U.S. Cyber Command: Waging War in the World's Fifth Battlespace* (Montreal: Centre for Research on Globalization, May 27 2010), http://www.globalresearch.ca/index.php?context=va&aid=19360

[3] R. Powell, *Nuclear Deterrence Theory: The Search for Credibility* (Cambridge: Cambridge University Press, 1990), p8

[4] A. Cohen, *Israel and the Bomb* (Columbia University Press, 1998)

[5] Department of Defense Directive 8500.01E, "Information Assurance (IA)", October 24, 2002 (certified current as of April 23, 2007), p3, http://www.dtic.mil/whs/directives/corres/pdf/850001p.pdf

[6] P. Knightley, *The second oldest profession: spies and spying in the twentieth century* (Pimlico, 2003)

[7] E. Patterson, *Just War Thinking: Morality and Pragmatism in the Struggle against Contemporary Threats* (Lexington Books, 2007), p63-69

[8] D. Albright, P. Brannan, and C. Walrond, "Did Stuxnet Take Out 1,000 Centrifuges at the Natanz Enrichment Plant? Preliminary Assessment," *Institute for Science and International Security (ISIS)*, http://isis-online.org/uploads/isis-reports/documents/stuxnet_FEP_22Dec2010.pdf

[9] E. Sterner, "Retaliatory Deterrence in Cyberspace," *Strategic Studies Quarterly*, Spring 2011

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

Artificial Intelligence in Cyber Defense

Enn Tyugu
R&D Branch
Cooperative Cyber Defense Center of Excellence (CCD COE)
and Estonian Academy of Sciences
Tallinn, Estonia
tyugu@ieee.org

Abstract- **The speed of processes and the amount of data to be used in defending the cyber space cannot be handled by humans without considerable automation. However, it is difficult to develop software with conventional fixed algorithms (hard-wired logic on decision making level) for effectively defending against the dynamically evolving attacks in networks. This situation can be handled by applying methods of artificial intelligence that provide flexibility and learning capability to software. This paper presents a brief survey of artificial intelligence applications in cyber defense (CD), and analyzes the prospects of enhancing the cyber defense capabilities by means of increasing the intelligence of the defense systems. After surveying the papers available about artificial intelligence applications in CD, we can conclude that useful applications already exist. They belong, first of all, to applications of artificial neural nets in perimeter defense and some other CD areas. From the other side – it has become obvious that many CD problems can be solved successfully only when methods of artificial intelligence are being used. For example, wide knowledge usage is necessary in decision making, and intelligent decision support is one of yet unsolved problems in CD.**

Keywords: applied artificial intelligence; intelligent cyber defense methods; neural nets in cyber defense; expert systems in cyber defense.

.

This paper expresses the author's ideas, and does not reflect an official position of CCD COE. The present work has been partially supported by the CCD COE and by the Centre of Excellence in Computer Science (EXCS)

I. INTRODUCTION

It is obvious that defense against intelligent cyber weapons can be achieved only by intelligent software, and events of the last two years have shown rapidly increasing intelligence of malware and cyber-weapons. Let us mention the Conficker worm for example. Some effects of Conficker on military and police networks in Europe have been cited in [1] as follows: "Intramar, the French Navy computer network, was infected with Conficker on 15 January 2009. The network was subsequently quarantined, forcing aircraft at several airbases to be grounded because their flight plans could not be downloaded. The United Kingdom Ministry of Defense reported that some of its major systems and desktops were infected. The virus has spread across administrative offices, NavyStar/N* desktops aboard various Royal Navy warships and Royal Navy submarines, and hospitals across the city of Sheffield reported infection of over 800 computers. On 2 February 2009, the Bundeswehr, the unified armed forces of the Federal Republic of Germany reported that about one hundred of their computers were infected. In January 2010, the Greater Manchester Police computer network was infected, leading to its disconnection for three days from the Police National Computer as a precautionary measure; during that time, officers had to ask other forces to run routine checks on vehicles and people."

Application of network centric warfare (NCW) makes cyber incidents especially dangerous, and changes in cyber defense are urgently required [2]. The new defense methods like dynamic setup of secured perimeters, comprehensive situation awareness, highly automated reaction on attacks in networks will require wide usage of artificial intelligence methods and knowledge-based tools.

Why has the role of intelligent software in cyber operations increased so rapidly? Looking closer at the cyber space, one can see the following answer. Artificial intelligence is needed, first of all, for rapid reaction to situations in Internet. One has to be able to handle large amount of information very fast in order to describe and analyze events that happen in cyber space and to make required decisions. The speed of processes and the amount of data to be used cannot be handled by humans without considerable automation. However, it is difficult to develop software with conventional fixed algorithms (hard-wired logic on decision making level) for effectively defending against the attacks in cyber space, because new threats appear constantly. Here is a place for artificial intelligence methods.

The second section of the present paper introduces artificial intelligence as a field of science and technology. In the third section we look at the existing artificial intelligence applications in cyber defense, grouped by the artificial intelligence techniques. The fourth section looks into the future and suggests new intelligent applications.

II. ABOUT ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) as a field of scientific research (also called machine intelligence in the beginning) is almost as old as electronic computers are. A possibility of building devices/software/systems more intelligent than human beings has been from the early days of AI "on the horizon". The problem is that the time horizon moves away when time passes. We have witnessed the solving of a number of intelligently hard problems by computers like playing good chess, for instance. During the early days of computing the chess playing was considered a benchmark showing a real intelligence. Even in seventies of the last century, when the computer chess was on the masters level, it seemed almost impossible to make a program that could beat the world champion. However, this happened sooner than expected. This had three reasons: increased computing power, development of a good search algorithm (that can be used in many applications beside chess, see the section on search below), and well organized knowledge bases that included all available chess knowledge (first of all, opening and end games). In essence, the chess problem could be solved because it was a specific intellectual problem belonging to so called *narrow AI*. A different case is translating from one language into another that requires *general AI*. In sixties of the last century, especially after N. Chomski's work in structural linguistics, it was expected that the natural language translation problem will be solved soon. It has not happened yet, although success is visible in some specific applications like, for instance, Google's AI linguistics. The reason is that this requires artificial general intelligence -- possessing of and ability to handle large amounts of knowledge in every field related to human activities.

It is generally accepted that AI can be considered in two ways: as a science aimed at trying to discover the essence of intelligence and developing generally intelligent machines, or as a science providing methods for solving complex problems that cannot be solved without applying some intelligence like, for instance, playing good chess or making right decisions based on large amounts of data. In the present paper we will take the second approach, advocate for applying specific AI methods to cyber defense problems, and will refer to the existing AI algorithms described in [3].

III. WHAT WE HAVE TODAY

After surveying the papers available about AI applications in CD, we are able to conclude that numerous useful applications already exist in this field. They belong, first of all, to applications of artificial neural nets in perimeter defense. On the other hand – it has become obvious that many more CD problems can be solved successfully only when AI methods are used. Wide knowledge usage is necessary in decision making, and the intelligent decision support is one of the yet unsolved problems in CD.

A large number of methods have been developed in the artificial intelligence field for solving hard problems that require intelligence from the human perspective. Some of these methods have reached a stage of maturity where precise algorithms exist that are based on these methods. Some methods have even become so widely known that they are not considered belonging to artificial intelligence any more, but have become a part of some application area, for instance, data mining algorithms that have emerged from the learning subfield of AI. It would be impossible to try to give more or less complete survey of all practically useful AI methods in a brief survey. Instead, we have grouped the methods and architectures in several categories: neural nets, expert systems, intelligent agents, search, machine learning, data mining and constraint solving. We outline these categories here, and we give references to the usage of respective methods in cyber defense. We are not going to discuss natural language understanding, robotics and computer vision which we consider specific applications of AI. Robots and computer vision have definitely impressive military applications, but we have not found anything specific to CD there.

A. Neural nets

Neural nets have a long history that begins with the invention of *perceptron* by Frank Rosenblatt in 1957 – an artificial neuron that has remained one of the most popular elements of neural nets [4]. Already a small number of perceptrons combined together can learn and solve interesting problems. But neural nets can consist of a large number of artificial neurons. Therefore neural nets provide a functionality of massively parallel learning and decision-making. Their most distinguished feature is the speed of operation. They are well suited for learning pattern recognition, for classification, for selection of responses to attacks [5] etc. They can be implemented either in hardware or in software.

Neural nets are well applicable in intrusion detection and intrusion prevention [6, 7, 8, 9, 10]. There have been proposals to use them in DoS detection [11], computer worm detection [12], spam detection [13], zombie detection [14], malware classification [15] and in forensic investigations [16].

A reason for the popularity of neural nets in cyber defense is their high speed, if implemented in hardware or used in graphic processors. There are new developments in the neural nets technology: third generation neural nets – spiking neural networks that mimic biological neurons more realistically, and provide more application opportunities. Good opportunities are provided by the usage of FPGA-s (field programmable gate arrays) that enable rapid development of neural nets and their adjustment to changing threats.

B. Expert systems

Expert systems are unquestionably the most widely used AI tools. An expert system is software for finding answers to questions in some application domain presented either by a user or by another software [17]. It can be directly used for

decision support, e.g. in medical diagnosis, in finances or in cyberspace. There is a great variety of expert systems from small technical diagnostic systems to very large and sophisticated hybrid systems for solving complex problems. Conceptually, an expert system includes a *knowledge base*, where expert knowledge about a specific application domain is stored. Besides the knowledge base, it includes an *inference engine* for deriving answers based on this knowledge and, possibly, additional knowledge about a situation. Empty knowledge base and inference engine are together called *expert system shel*l -- it must be filled with knowledge, before it can be used. Expert system shell must be supported by software for adding knowledge in the knowledge base, and it can be extended with programs for user interactions, and with other programs that may be used in hybrid expert systems. Developing an expert system means, first, selection/adaptation of an expert system shell and, second, acquiring expert knowledge and filling the knowledge base with the knowledge. The second step is by far more complicated and time consuming than the first.

There are many tools for developing expert systems. In general, a tool includes an expert system shell and has also a functionality for adding knowledge to the knowledge repository. Expert systems can have extra functionality for simulation [5], for making calculations etc. There are many different knowledge representation forms in expert systems, the most common is a rule-based representation. But the usefulness of an expert system depends mainly on the quality of knowledge in the expert system's knowledge base, and not so much on the internal form of the knowledge representation. This leads one to the *knowledge acquisition problem* that is crucial in developing real applications.

Example of a CD expert system is one for security planning [18]. This expert system facilitates considerably selection of security measures, and provides guidance for optimal usage of limited resources. There are early works on using expert systems in intrusion detection [19, 20].

C. Intelligent agents

Intelligent agents are software components that possess some features of intelligent behavior that makes them special: proactiveness, understanding of an agent communication language (ACL), reactivity (ability to make some decisions and to act). They may have a planning ability, mobility and reflection ability. In the software engineering community, there is a concept of software agents where they are considered to be objects that are at least proactive and have the ability to use the agent communication language. Comparing agents and objects, one can say that objects may be passive, and they do not have to understand any language (although they accept messages with well-defined syntax.)

Using intelligent agents in defense against DDoS has been described in [21] and [22], where simulation shows that cooperating agents can effectively defend against DDoS attacks. After solving some legal [23] 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 unaccessible for adversaries. This will require cooperation with ISP-s. Multi-agent tools can provide more complete operational picture of the cyber space, for instance, a hybrid multi-agent and neural network-based intrusion detection method has been proposed in [24]. Agent-based distributed intrusion detection is described in [25].

D. Search

Search is a universal method of problem solving that can be applied in all cases when no other methods of problem solving are applicable. People apply search in their everyday life constantly, without paying attention to it. Very little must be known in order to apply some general search algorithm in the formal setting of the search problem: one has to be able to generate candidates of solutions, and a procedure (formally a predicate) must be available for deciding whether a proposed candidate satisfies the requirements for a solution. However, if additional knowledge can be exploited to guide the search, then the efficiency of search can be drastically improved. Search is present in some form almost in every intelligent program, and its efficiency is often critical to the performance of the whole program.

A great variety of search methods have been developed which take into account the specific knowledge about particular search problems. Although many search methods have been developed in AI, and they are widely used in many programs, it is seldom considered as the usage of AI. For example, in [26] and [27] dynamic programming is essentially used in solving optimal security problems, the search is hidden in the software and it is not not visible as an AI application. Search on and-or trees, αβ-search, minimax search and stochastic search are widely used in games software, and they are useful in decision-making for cyber defense. The αβ-search algorithm, originally developed for computer chess, is an implementation of a generally useful idea of "divide an conquer" in problem solving, and especially in decision making when two adversaries are choosing their best possible actions. It uses the estimates of minimally guaranteed win and maximally possible loss. This enables one often to ignore large amount of options and considerably to speed up the search.

E. Learning

Learning is improving a knowledge system by extending or rearranging its knowledge base or by improving the inference engine [28]. This is one of the most interesting problems of artificial intelligence that is under intensive investigation. Machine learning comprises computational methods for acquiring new knowledge, new skills and new ways to organize existing knowledge. Problems of learning vary greatly by their complexity from simple *parametric learning* which means learning values of some parameters, to complicated forms of *symbolic learning*, for

example, learning of concepts, grammars, functions, even learning of behavior [29].

AI provides methods for both -- *supervised learning* (learning with a teacher) as well as *unsupervised learning*. The latter is especially useful in the case of presence of large amount of data, and this is common in cyber defense where large logs can be collected. Data mining has originally grown out of unsupervised learning in AI. Unsupervised learning can be a functionality of neural nets, in particular, of self-organizing maps [30, 10, 16, 31].

A distinguished class of learning methods is constituted by parallel learning algorithms that are suitable for execution on parallel hardware. These learning methods are represented by genetic algorithms and neural nets. Genetic algorithms and fuzzy logic has been, for instance, used in threat detection systems described in [32].

F. Constraint solving

Constraint solving or constraint satisfaction is a technique developed in AI for finding solutions for problems that are presented by giving a set of constraints on the solution, e.g. logical statements, tables, equations, inequalities etc. [3, 33]. A solution of a problem is a collection (a tuple) of values that satisfy all constraints. Actually, there are many different constraint solving techniques, depending on the nature of constraints (for example, constraints on finite sets, functional constraints, rational trees). On a very abstract level, almost any problem can be presented as a constraint satisfaction problem. In particular, many planning problems can be presented as constraint satisfaction problems. These problems are difficult to solve because of large amount of search needed in general. All constraint solving methods are aimed at restricting the search by taking into account specific information about the particular class of problems. Constraint solving can be used in situation analysis and decision support in combination with logic programming [34, 35].

IV. CHALLENGES IN INTELLIGENT CYBER DEFENSE

When planning the future research, development and application of AI methods in CD, one has to distinguish between the immediate goals and long-term perspectives. There are numerous AI methods immediately applicable in CD, and there are immediate CD problems that require more intelligent solutions than have been implemented at present. Until now we have discussed these existing immediate applications.

In the future, one can see promising perspectives of the application of completely new principles of knowledge handling in situation management and decision making. These principles include introduction of a *modular and hierarchical knowledge architecture* in the decision making software. This kind of architecture

has been proposed in [36]. A challenging application area is the knowledge management for net centric warfare [37]. Only automated knowledge management can guarantee rapid situation assessment that gives a decision superiority to leaders and decision makers on any C2 level. As an example, the paper [36] describes an idea of the hierarchical and modular knowledge architecture in the Joint Command and Control Information System of the Bundeswehr.

Expert systems are already being used in many applications, sometimes hidden inside an application, like in the security measures planning software [26]. However, expert systems can get wider application, if large knowledge bases will be developed. This will require considerable investment in knowledge acquisition, and development of large modular knowledge bases. Also further development of the expert system technology will be needed: modularity must be introduced in the expert system tools, and hierarchical knowledge bases must be used.

Considering a more distant future -- at least some decades ahead, perhaps we should not restrict us to the "narrow AI". Some people are convinced that the grand goal of the AI -- development of artificial general intelligence -- AGI can be reached in the middle of the present century. The first conference on AGI was held in 2008 at the University of Memphis. The Singularity Institute for Artificial Intelligence (SIAI), founded in 2000, warns researchers of a danger that exponentially faster development of intelligence in computers may occur. This development may lead to Singularity, described in [38] as follows: "The Singularity is the technological creation of smarter-than-human intelligence. There are several technologies that are often mentioned as heading in this direction. The most commonly mentioned is probably Artificial Intelligence, but there are others ... -- several different technologies which, if they reached a threshold level of sophistication, would enable the creation of smarter-than-human intelligence. ... A future that contains smarter-than-human minds is genuinely different in a way that goes beyond the usual visions of a future filled with bigger and better gadgets." A futurist Ray Kurtzwell has extrapolated the development to come up with Singularity in 2045 [39]. One need not to believe in the Singularity threat, but the rapid development of information technology will definitely enable one to build considerably better intelligence into software in coming years. (Consider the recent impressive performance of IBM-s Watson program [40].) Independently of whether the AGI is available or Singularity comes, it is crucial to have the ability to use better AI in cyber defense than the offenders have it.

V. CONCLUSIONS

In the present situation of rapidly growing intelligence of malware and sophistication of cyber attacks, it is unavoidable to develop intelligent cyber defense methods. The experience in DDoS mitigation has shown that even a defense against large-scale attacks can be successful with rather limited resources when intelligent methods are used.

An analysis of publications shows that the AI results most widely applicable in CD are provided by the research in artificial neural nets. Applications of neural nets will continue in CD. There is also an urgent need for application of intelligent cyber defense methods in several areas where neural nets are not the most suitable technology. These areas are decision support, situation awareness and knowledge management. Expert system technology is the most promising in this case.

It is not clear how rapid development of general artificial intelligence is ahead, but a threat exists that a new level of artificial intelligence may be used by the attackers, as soon as it becomes available. Obviously, the new developments in knowledge understanding, representation and handling [41, 42, 43] as well in machine learning will greatly enhance the cyber defense capability of systems that will use them.

REFERENCES

[1] http://en.wikipedia.org/wiki/Conficker

[2] R. A. Poell, P. C. Szklrz. R3 – Getting the Right Information to the Right People, Right in Time. Exploiting the NATO NEC. In: M.- Amanovicz. Comcepts and Implementations for Innovative Military Communications and Information Technologies. Military University of Technology Publisher, Warsaw, 2010, 23 – 31.

[3] E. Tyugu. Algorithms and Architectures of Artificial Intelligence. IOS Press. 2007.

[4] F. Rosenblatt. The Perceptron -- a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory, 1957.

[5] G. Klein, A. Ojamaa, P. Grigorenko, M. Jahnke, E. Tyugu. Enhancing Response Selection in Impact Estimation Approaches. Military Communications and Information Systems Conference (MCC), Wroclaw, Poland, 2010.

[6] J. Bai, Y. Wu, G. Wang, S. X. Yang, and W. Qiu, "A novel intrusion detection model based on multi-layer self-organizing maps and principal component analysis," in Advances in Neural Networks - ISNN 2006, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, May 2006, pp. 255–260.

[7] F. Barika, K. Hadjar, and N. El-Kadhi, "Artificial neural network for mobile IDS solution," in Security and Management, 2009, pp. 271–277.

[8] D. A. Bitter, T. Elizondo, Watson. Application of Artificial Neural Networks and Related Techniques to Intrusion Detection. WCCI 2010 IEEE World Congress on Computational Intelligence. July, 18-23, 2010 - CCIB, Barcelona, Spain, 2010, pp. 949 – 954.

[9] R.-I. Chang, L.-B. Lai, W. D. Su, J. C. Wang, and J.-S. Kouh, "Intrusion detection by backpropagation neural networks with sample-query and attribute-query," International Journal of Computational Intelligence Research, vol. 3, no. 1, 2007, pp. 6–10.

[10] L. DeLooze, Attack Characterization and Intrusion Detection using an Ensemble of Self-Organizing Maps, Proceedings of the IEEE Workshop on Information Assurance United States Military Academy, West Point, NY, 2006.

[11] B. Iftikhar, A. S. Alghamdi, "Application of artificial neural network in detection of dos attacks," in SIN '09: Proceedings of the 2nd international conference on Security of information and networks. New York, NY, USA: ACM, 2009, pp. 229–234.

[12] D. Stopel, Z. Boger, R. Moskovitch, Y. Shahar, and Y. Elovici, "Application of artificial neural networks techniques to computer worm detection," in International Joint Conference on Neural Networks (IJCNN), 2006, pp. 2362–2369.

[13] C.-H. Wu, "Behavior-based spam detection using a hybrid method of rule-based techniques and neural networks," Expert Systems with Applications, vol. 36, no. 3, Part 1, 2009, pp. 4321–4330.

[14] P. Salvador et al. Framework for Zombie Detection Using Neural Networks. In: Fourth International Conference on Internet Monitoring and Protection ICIMP-09, 2009.

[15] M. Shankarapani, K. Kancherla, S. Ramammoorthy, R. Movva, and S. Mukkamala. Kernel Machines for Malware Classification and Similarity Analysis. WCCI 2010 IEEE World Congress on Computational Intelligence. Barcelona, Spain, 2010, pp. 2504 – 2509.

[16] B. Fei, J. Eloff, MS Olivier, H. Venter. The use of self-organizing maps of anomalous behavior detection in a digital investigation. Forensic Science International, v. 162, 2006,pp. 33-37.

[17] http://en.wikipedia.org/wiki/Expert_system. Expert System. Wikipedia.

[18] J. Kivimaa, A. Ojamaa, E. Tyugu. Graded Security Expert System. Lecture Notes in Computer Science, v. 5508. Springer, 2009, 279-286.

[19] D. Anderson, T. Frivold, A. Valdes. Next-generation intrusion detection expert system (NIDES). Technical Report SRI-CSL-95-07, SRI International, Computer Science Lab (1995).

[20] TF. Lunt, R. Jagannathan. A Prototype Real-Time Intrusion-Detection Expert System. Proc. IEEE Symposium on Security and Privacy, 1988, p. 59.

[21] I. Kotenko, A. Ulanov. Multi-Agent Framework fo Simulation of Adaptive Cooperative Defense Against Internet Attacks. In: International Workshop on Autonomous Intelligent Systems: Agents and Data Mining. LNCS, Springer, v. 4476.

[22] I. Kotenko, A. Konovalov, A.Shorov. Agent-Based modeling and Simulation of Botnets and Botnet Defence. In: C. Czosseck, K. Podins (eds.). Proc. Conference on Cyber Conflict. CCD COE Publications, Tallinn, Estonia, 2010.

[23] 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.

[24] E. Herrero, M. Corchado, A. Pellicer, A. Abraham, "Hybrid multi agent-neural network intrusion detection with mobile visualization," Innovations in Hybrid Intelligent Systems, vol. 44, 2007, pp. 320–328.

[25] V. Chatzigiannakis, G. Androulidakis, B. Maglaris. A Distributed Intrusion Detection Prototype Using Security Agents. HP OpenView University Association, 2004.

[26] J. Kivimaa, A. Ojamaa, E. Tyugu. Pareto-Optimal Situation Analysis for Selection of Security Measures. Proc. MilCom, 2008.

[27] J. Kivimaa, A. Ojamaa, E. Tyugu. Managing Evolving Security Situations. MILCOM 2009: Unclassified Proceedings, Boston, MA. Piscataway, NJ: IEEE, 2009, pp. 1 - 7.

[28] P. Norvig, S. Russell. Artificial Intelligence: Modern Approach. Prentice Hall, 2000.

[29] AK. Ghosh, C. Michael, M. Schatz. A Real-Time Intrusion Detection System Based on Learning Program Behavior. Proceedings of the Third International Workshop on Recent Advances in Intrusion Detection, 2000, pp.93-109.

[30] J. Bai, Y. Wu, G. Wang, S. X. Yang, and W. Qiu, A novel intrusion detection model based on multi-layer self-organizing maps and principal component analysis, in Advances in Neural Networks. Lecture Notes in Computer Science. Springer, 2006, pp. 255–260.

[31] V. K. Pachghare, P. Kulkarni, D. M. Nikam. Intrusion Detection System using Self Organizing Maps. Proc. International Conference on Intelligent Agent & Multimedia-Agent Systems, IAMA 2009.

[32] R. Hosseini, J. Dehmeshki, S. Barman, M. Mazinani, S. Qanadli . A Genetic Type-2 Fuzzy Logic System for Pattern Recognition in Computer Aided Detection Systems.

IEEE World Congress on Computational Intelligence. Barcelona, Spain, 2010, pp. 215 – 221.

[33] B. Mayoh, E. Tyugu, J. Penjam. Constraint Programming. NATO ASI Series, v. 131, Springer Verlag. 1994.

[34] I. Bratko. PROLOG Programming for Artificial Intelligence. Addison-Wesley, 2001 (third edition).

[35] Xinming Ou. A logic-programming approach to network security analysis. PhD Thesis, Princeton University, 2005.

[36] U. Kaster, B. Kuhiber. Information and Knowledge Management in C2 Systems – The Gap Between Theory and Practice is not all that big. In: M.- Amanovicz. Comcepts and Implementations for Innovative Military Communications and Information Technologies. Military University of Technology Publisher, Warsaw, 2010, pp. 98 – 107.

[37] J. Kaster. Combined Knowledge Management and Workflow Management in C2 Systems – a user centered approach. Fraunhofer Institute for Communication, Information Processing and Ergonomics. Report ID # 197, 2009.

[38] http://singinst.org/overview/whatisthesingularity/

[39] R. Kurtzwell. The Singularity is Near. Viking Adult. 2005.

[40] http://www.ted.com/webcast/archive/event/ibmwatson

[41] M. Chmielewski. Ontology Applications for Achieving Situation Awareness in Military Decision Support Systems. LNAI/LNCS, Proc, ICCCI 2009, Wroclaw, 2009.

[42] P. Lorents, E. Tyugu Lattices of knowledge systems. Proc. International Conference on Artificial Intelligence Proc. WORLDCOMP'09: IC-AI´2009, Las Vegas, CSREA Press, July 2009.

[43] U. Schade, M. R. Hieb. A Battle Management Language for Orders, Requests and Reports. In: 2007 Spring Simulatin Ineroperability Workshop. Norfolk, USA, 2006

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

On the Arms Race Around Botnets – Setting Up and Taking Down Botnets

Christian Czosseck
Cooperative Cyber Defence Centre of Excellence
Tallinn, Estonia
christian.czosseck@ccdcoe.org

Gabriel Klein
Fraunhofer FKIE
Wachtberg, Germany
gabriel.klein@fkie.fraunhofer.de

Felix Leder
Institute of Computer Science 4
University of Bonn
Germany
leder@cs.uni-bonn.de

Abstract—**Botnets are a well-recognized and persistent threat to all users of the Internet. Since the first specimens were seen two decades ago, botnets have developed form a subject of curiosity to highly sophisticated instruments for illegally earning money. In parallel, an underground economy has developed which creates hundreds of millions of euros per year in revenue with spamming, information theft, blackmailing or scare-ware. Botnets have become a high-value investment for their operators that need to be protected from law enforcement agencies or the anti-botnet community. Security researchers and companies trying to keep them within bounds are facing the very latest in spreading and defense techniques. Hundreds of thousands of new malware samples per month pose an immense challenge for AV companies. Specialized countermeasures against botnets have evolved along with botnet technology, trying to bring them down by targeting the root of every botnet: its command-and-control structure. This leads to an ongoing arms race between botnet developers and their operators vs. security experts. So far the former have the upper hand.**

Based on the analysis of multiple botnet takedowns and the in-depth investigation of various botnet architectures conducted by the authors, this paper provides an analysis of the efforts needed to acquire and set up a botnet. This is followed by a comparison of selected significant botnet countermeasures, which are discussed with regard to their required resources. Legal and ethical issues are also addressed, while a more thorough discussion of these will be left for future work.

Keywords-IT security; botnet; malware; infection; disinfection; botnet setup; botnet takedown; tactical takedown;

I. INTRODUCTION: CURRENT STATE OF THE ARMS RACE

Botnets are networks of computers infected with malicious software (malware), remotely controlled by so-called bot herders. The infected machines within this botnet (a.k.a. bot or zombie) are regularly abused to perform mostly criminal activities without the knowledge of their owners. This includes but is not limited to sending spam, conducting distributed denial-of-service (DDoS) attacks or harvesting sensitive data such as credit card credentials. Beyond credit card fraud, extortion schemes can also be observed with threats of large-scale DDoS attacks unless payments are made. All this leads to steadily increasing financial damage and cyber crime's yearly income surpassing the global drugs revenue [1]. As a latest trend, botnets play an increasing role in politically motivated attacks against public and private institutions, sometimes threatening entire countries [2]. Behind this is a well-developed underground market, on which botnet technology and associated services are sold to everyone at rather low prices [15].

Anti-virus (AV) companies as the natural enemy of malware are constantly trying to keep up with the growing threat, developing a variety of products to protect computers from being infected. Unfortunately, malware authors are often one step ahead because of the reactive defense provided by AV software. If newly developed malware is released, even up-to-date anti-virus detectors are often not likely to detect it [3]. Some AV software detects less than 10 % of new samples within the first 24 hours of their occurrence. Often manual analysis of new malware samples is required because automatic approaches are limited in their capabilities. There is a general consensus in the AV industry that current solutions are neither scalable nor sustainable enough.

Acknowledging the fact that malware spreading cannot be stopped or slowed down significantly, other countermeasures directly attacking established botnets have been developed.

To receive or pass on commands, the individual parts of these botnets need to communicate with each other and with their so-called command-and-control (C&C) servers. The method according to which this communication takes place defines the topology of the botnet. So far three different ones have been established: centralized topologies with few C&C servers, decentralized topologies based on peer-to-peer (P2P) protocols, and semi-flexible topologies often realized by fluxy domain registrations.

If connectivity between bots and C&C servers is established, different communication protocols like HTTP or IRC are commonly used. A more in-depth discussion of technical botnet issues can be found in [4] and [5]. All these technical aspects provide entry vectors for targeted counter measures against botnets.

This paper provides a comprehensive overview of the resources needed in this arms race between bot herders and botnet hunters. Based on analyses of recent botnet investigations and experience from conducted takedowns, the most common countermeasures are presented and analyzed.

The paper discusses the countermeasures with regard to their required resources, namely *required skills*, *monetary costs* and *time* as well as the *likelihood of a*

successful takedown. Legal and ethical aspects are also addressed. The discussion is based on a simple taxonomy of botnets presented in Section 2, grouping botnets into three broad groups. In Section 3, we discuss the efforts needed to set up a botnet for each of the introduced categories. This is followed by an in-depth discussion on required resources for the most common botnet countermeasures in Section 4. We conclude with a summary and an outlook on future developments.

II. BOTNET EVOLUTION

Since the world fist encountered a computer virus called Brain back in 1984, malware has developed from a proof of creativity to a highly sophisticated instrument usable for various tasks, nowadays aiming mainly for earning money in a criminal way.

Botnets themselves evolved from the idea of a massive remote control for administrative tasks to a flexible type of malware encompassing the most successful spreading and hiding techniques.

Botnets evolved and became more professional over time. This is reflected in their capabilities, but also in the skills needed for their creation. Nowadays, experienced and knowledgeable malware developers are needed to create botnets which are hard to detect or to mitigate.

This paper introduces three broad categories of botnets reflecting the major evolutional steps over the past decades. The discussion on setting up or taking over/down botnets is structured according to these categories.

A. Open-Source Botnets

The first category is formed by botnets that were either developed open-source, were made freely available later on, or are easy to find. This marks the very beginning of malicious botnet development, where botnets and malware in general was often written for (often still illegal) fun or out of competition between “geeks”. Monetary aspects were hardly a driving force.

Well-known representatives are botnets like AgoBot, SDBot or RBot [22, 23]. While quite old, they are still in use and are sometimes developed further by single groups adding new exploits or functionality. These new exploits are developed individually or obtained from other sources like the exploit framework Metasploit [6]. For this category we assume that most of the code base is freely available for both malware developers and AV companies. They are easy to set up, typically only requiring the botherder to make some changes in provided configuration files and compiling the code.

An alternative way of operation is the development of a closed-source botnet (which might be a fork of or inspired by an existing open-source botnet), adding well developed open-source components to the code base. Popular examples for open-source components integrated into closed-source botnets are cryptographic and compression routines. Waledac used the OpenSSL library [4, 7], for both RSA and AES, Conficker included the official MIT implementation of the MD6 hash algorithm [8], and Storm made use of the zlib [9] compression library [24]. This

provides malware developers with reliable, well tested standard routines, reducing their efforts. If a particular botnet is a fork of one of the older open-source botnets mentioned above, we consider it to fall under this category. Otherwise the botnets belong to one of the following categories.

B. *Construction Kit-based Botnets*

Over time, botnets developed into an effective tool to illicitly earn money. With AV and other security companies putting more efforts into fighting botnets, botnet developers improved resistance to countermeasures. However, they invented an increasing number of new features for generating money, e. g. by harvesting financial credentials or other valuable information and subsequently selling them later. Botnet developers started to understand the value of their creations and that not everybody is able to develop sophisticated botnets, thus raising the value of well-developed ones. Out of this a business model developed in an underground scene, where botnet developers started to sell botnet software. To an increasing extent, this is offered together with patch services, infection guarantees and/or hosting services. Botnets became available as so-called constructions kits, enabling everyone to configure and create their own botnet in a point-and-click fashion.

This category covers all botnets fitting this description. They are assumed to be well maintained, regularly updated, and coming with the ability to add new functionality (add-ons), maybe even by third parties. Many of them are sold including support for the buyer via ICQ or other digital media. These botnets are normally developed as closed source, using state-of-the-art methods, software development processes and quality assurance methods [10]. To protect the botnet software, licensing schemes and code protection software such as VMProtect [11], commonly encountered in legitimate software products, are used to control the distribution of their products. This makes analyzing or stealing the botnet's source code hard for competitors and AV companies.

Prominent examples of botnets that are sold as construction kits are ZeuS and its presumable successor, SpyEye, both targeting financial data. In the case of ZeuS, the C&C server is provided based on open-source components written in PHP. Prices usually range from a few hundred to several thousand USD depending on the feature set [15].

C. *Specialized Botnets*

The last category this paper introduces covers all botnets, which were developed with a very specific target or functionality in mind. While most of the attributes of the second category still apply, specialized botnets are highly professionally developed, combining advanced expertise in exploit development (e. g. by usage of 0-day exploits), careful software engineering considering latest countermeasures, and sometimes even combine cross-domain knowledge or intelligence of the target. Monetary gains as a driving force might but do not need to be present. Espionage and sabotage are other motivations for this advanced persistent threat (APT).

Examples for this category are Ghostnet [12], which aimed at political espionage in China-critical communities, or Stuxnet, which was developed to target Windows-based supervisory control and data acquisition (SCADA) systems and is assumed to be an instrument of information gathering and sabotage of the Iranian nuclear program [13]. Night dragon is a botnet spying mainly on petrol and gas companies [25]. Another example is Conficker, which, while actively developed to be impervious to latest countermeasures and widely spread, still has not shown any active functional payload.

III. SETTING UP A BOTNET

In 2008, spammers alone earned an estimated 780 million USD [14] and there is an upward trend to these numbers. With this ever increasing amount of money to be made by operating or renting out botnets, an increasing professionalization in the domain can be observed [26]. Structures similar to free market (sub-) economies are emerging where prices and the availability of products and services are regulated by demand. There are even marketing campaigns on underground forums promoting certain products. Taking these issues into account, what are the resources that remain to be expended for setting up and deploying a botnet? In this section we will discuss these resources in the context of the botnet categories introduced in the previous section.

A. Finances

As the development of open-source botnets is community-driven, no direct monetary cost is involved. Depending on the situation, software developers might need to be paid.

The prices of construction kit botnets vary; entry-level ZeuS kits can be purchased for 3,000-4,000 USD, whereas more advanced kits can cost more than 10,000 USD [15]. Additionally, appropriate infection kits can be bought from 100 to more than 1,000 USD [16]. A range of companies exist that provide “all-inclusive” packages for botnet construction, propagation with exploit kits, as well as command-and-control server hosting and maintenance.

Where specialized botnets are concerned, especially skilled and trusted developers are required. Components are typically self-developed and seldom purchased. The required trust and skill level makes these types of botnets more expensive than extensions to open-source botnets. In case of sabotage and for reliable development, test environments have to be bought, set up, and maintained [13].

B. Development Skills

There are three aspects to be considered when developing a botnet: the infection of the target machine, the botnet binary that is executed on the target machine after infection, and the C&C infrastructure. Different development skills are required for each aspect. Resource requirements for infection are fairly similar for the different classes of botnets and are discussed in a subsequent section.

To configure and install the bot component of an open-source botnet, the user needs a basic understanding of source code configuration and needs to be able to use a compiler. A non-technical user can acquire these skills in a short amount of time. However, a risk in this case is the unknown programming quality of the malware.
Compared to this, the configuration and setup of construction kit botnets is almost negligible. These kits are for sale and designed with user-friendliness in mind. The entire process takes only a few clicks of the mouse. Configuration is accomplished easily by adapting existing configuration files or purchasing ready-made ones. Most kits come with a standard set of system manipulations.
With specialized botnets, the greatest difficulty lies in the amalgamation of cross-domain knowledge, . This does not usually apply to botnets in the other groups. Specialized botnets have highly customized goals, e. g. espionage or sabotage. Exploiting weaknesses and optimizing malware for execution in systems in these environments requires a high degree of immersion in that context. An example of this is Stuxnet. Here, a detailed familiarity with very specific industrial control systems was required. The actual technical skills are comparable to those required for open-source botnets, although in most cases there is no software base to build upon so extensive development effort is needed. An additional difficulty is that community support cannot be relied on here.
Where the development of C&C infrastructure is concerned, construction kit-based botnets require the least effort of the three classes. In principle, setting up a C&C server is identical to setting up any other content-management system. For a more in-depth discussion of C&C infrastructures, please refer to [4]. Protecting the C&C server against takedown attempts by authorities and security researchers is more challenging, but often all-inclusive bundles are offered that include setup, support and bulletproof hosting of the C&C server. For open-source and specialized botnets, these activities have to be performed manually.

C. *Defensive Skills*

To protect their software from reverse engineering and analysis, malware authors increasingly employ defensive measures on a technical level.
An often-used mechanism is encryption, both of the communication with the C&C server and of the malware binary itself. Circumvention of the former is always possible. This is an imminent weakness in botnets using encrypted communication because the encryption keys either need to be included in the binary or can be observed when processed in the binary during runtime.
Obfuscation is a technique for hiding that different malware samples belong to the same botnet and to complicate detailed analysis of the internals. A recent trend is so-called server-side polymorphism. Here, the server from which a newly infected machine retrieves the actual malware encodes the binary differently for every client. This can include differences in the encryption routine, encryption keys, etc. The result is that binaries from two different infected hosts have nothing in common at first glance.

Already existing malware can be precisely immunized against certain AV products or analysis tools. This can easily be done manually because of Web sites such as VirusTotal, a meta-anti-virus tool that allows the online scanning of malware samples with multiple AV solutions. Malware can also be hardened automatically using third-party tools.

By implementing blacklists of IP addresses of known honeypot or other analysis systems, malware developers can explicitly avoid infecting malware analysis systems. Knowledge about such systems can be gathered in a variety of ways. A Web site called AV Tracker [18] contains a comprehensive list of sandboxes, Honeypots and other analysis systems operated by the AV industry and malware researchers world-wide. Going one step further, ZeuS operators have been observed to set up a honeypot-like system to analyze and provide further information about researchers trying to infiltrate its administrative interface [19].

In the case of open-source botnets, the employed defensive measures are hardly sophisticated and are mostly self-developed. Sometimes, adaptations of standard mechanisms can be observed. Botnets made with construction kits typically either have the defensive measures integrated into the construction kit or make use of so-called defensive kits. This modular technique allows the integration of arbitrary defensive measures into the construction kit just before the malware binary is compiled. Depending on the sophistication of these defensive kits, they are either freely available or need to be purchased. Defensive measures for specialized botnets are normally a mixture of standard techniques along with custom-built developments that ensure the stealth properties of the malware binary.

Because security researchers actively study and circumvent these defensive mechanisms, the result is a constant arms race in which botnet operators and developers continually develop new and more advanced mechanisms which are then analyzed and bypassed by the security industry.

D. Deployment

Originally, malware spread by exploiting server services via portable disks, nowadays often USB sticks.

A new trend is the increasing exploitation of vulnerabilities in client-side applications. These are often ubiquitous on user desktops and thus an easy target. Examples for these kinds of applications are Adobe's Portable Document Format (PDF) reader and Flash, Microsoft Office programs, or Web browsers. The latter can be exploited by so-called drive-by downloads on infected Web sites. These Web sites can be both legitimate sites hacked by criminals, or sites especially set up for the express purpose of infection. In the latter case, mass spam e-mails containing links to these sites lured users to these sites. According to the Websense 2010 threat report, 79.9 % of Web sites with malicious code were compromised legitimate sites [20].

Lately, there has been an increase in so-called targeted attacks which contain a social engineering component. Detailed background information is gathered on the intended targets and personalized messages are sent to the victims, either via e-

mail or through social networking sites. By exploiting information about the target's current personal or professional situation (e. g. hobbies or work-related activities), the target can be tempted to open either infected attached files or visit suggested Web sites.
When open-source botnets are employed, the infection routines are generally self-developed or developed and shared within the community. When specialized botnets are used, the situation is similar, but for different reasons. Here, secrecy and often the environment in which the botnet is operating necessitate own developments. In construction kit-based botnets, infection vectors are usually supplied in the form of the already mentioned exploit kits.
The time required for the infection of hosts is difficult to estimate for all three classes of botnets. This also depends on the definition of "a sufficient number" of nodes which can be different for different purposes. When botnets are created for renting or selling them to third-parties [27], a common size of 10,000 hosts is bundled. For open-source and kit-based botnets, 10,000 hosts can often be infected within several hours to several days. In seldom cases, this can take more than a week. Specialized botnets often do not have the target of maximum infection speed as their purpose may not be financial gain but rather the accomplishment of long-term goals such as espionage. Thus, infection speed may not be of the utmost importance.

IV. Resources Required for Tactical Countermeasures

Most of the common defensive techniques, such as firewalls, IDS, or anti-virus solutions, act on a local level. The locality is a problem when multiple targets are attacked that are managed by different entities, e. g. organizations with independent but cooperating branches. In addition, local measures can usually not prevent specific types of attacks, like DDoS attacks. A more sustainable and reliable way to counter such attacks is to conduct tactical countermeasures against the originating botnet.
In the following we will discuss the resources required to conduct different countermeasures that have the potential to take down the whole botnet. Two major types of countermeasures are considered. The first is classical countermeasures, which are rather moderate in their implications, but are very limited because of their dependence on the cooperation of other organizations. The second type is more aggressive countermeasures with global consequences which can be conducted by a single organization and are therefore, more suitable for a tactical take-down.
Each of the presented countermeasures is discussed with regard to the resources money, human resources and skill-level, cross-domain expertise required by those, time for conducting the countermeasure, sustainability, and possible legal or ethical constraints. Since many factors influence the different resources, no hard numbers are given but rather important relationships and estimations are explained.

A. Classical Countermeasures

1) C&C Server Takedown

If the location of a C&C server has been determined, it can theoretically be shut down or disconnected. This can be made difficult if redundant infrastructures spread multiple instances of the server all over the world, in particular hosting them with different providers. In addition to the main C&C endpoint(s), backup channels have to be identified, if the takedown is to be sustainable. If this has been achieved, sustainability is usually very high, especially for kit-based botnets for which details about the infrastructure are either freely available or can be purchased from security companies or malware intelligence (e. g. [28]). The same is true for open-source botnets, as source code analysis can easily reveal structural information. Specialized botnets, due to their stealthy nature, require significant effort in malware dissection by reverse engineering and forensics along with time and money to identify C&C endpoints and backup channels. .

Besides the required skills and money, cross-domain challenges like organizing cooperation with Internet service providers and local law enforcement authorities need to be faced. In an ideal case, the required time is in the vicinity of one hour. However, if lengthy analysis is needed and actions have to be coordinated with law enforcement in different countries, the entire process could take several months, if it is possible at all.

Legal constraints in some countries prohibit or complicate the takedown of C&C servers, enabling so-called bulletproof hosting requiring law enforcement intervention. In some countries, authorities and ISPs are reluctant to cooperate with security researchers or other security authorities. This is well-known and taken advantage of by botnet operators. Some ISPs notify customers if a site is about to be taken down and botnet operators can move the C&C server to another provider or a different country entirely.

2) DNS-based Countermeasures

If the C&C infrastructure of the botnet is based on DNS, then a classical countermeasure is deregistration of those domains required by the botnet. This has to be done in cooperation with the respective DNS registrars and was successful in several cases. A requirement for this countermeasure to be sustainable is that the botnet's C&C infrastructure relies solely on DNS mechanisms. If this requirement is met, DNS countermeasures are independent of the class of botnet, although C&C mechanisms tend to be more sophisticated in kit-based (Twitter-based selection of C&C server names in Torpig) and specialized (Kraken, Conficker) botnets.

Where money, skills and cross-domain knowledge is concerned, the main organizational challenge is the cooperation with the DNS registrars. These companies have no immediate benefit from such cooperation and often do so mainly because of the publicity effect. National and international law enforcement agencies also need to be coordinated with as there are legal issues to be considered. In the majority of cases, a court warrant needs to be obtained.

The time needed for this countermeasure to come into effect is affected by both the duration of the legal proceedings, i. e. to obtain the court warrant, and the time it takes for the DNS settings to be propagated to other servers. The latter can take from several minutes to several days, depending on the DNS time-to-live settings.
Already connected computers are not affected by this countermeasure; only newly connecting hosts performing a lookup receive the false information. Thus, the size of the botnet steadily decreases. The sustainability is very high.

B. *Proactive Countermeasures*

Beside the classical countermeasure, there are also more effective proactive countermeasures.

1) Response DDoS

If the locations of the C&C endpoints are known, a possibility is to launch a counter-DDoS attack to disable these endpoints. This is only possible if there is a single or limit number of C&C servers and would not work in a botnet relying on P2P infrastructure. A requirement for this is the availability of one or more machines for creating the traffic.
Financial resources are needed for the setup and operation of the traffic creation machines. This could, for example, be done by renting a competing botnet. According to [21], a DDoS botnet can be rented from 200 USD per 24 hours or 500 USD per month. Experiments conducted by an unnamed source have shown that a range of C&C servers can almost be shut down by DoS attacks from a single machine. This countermeasure is generally independent of the category of botnet being attacked. However, to determine the botnet's operating parameters, especially its C&C endpoints, can require extensive analysis. The resources in terms of skills, cross-domain activities and money required for this are comparable to those of the C&C server takedown described earlier.
The application of a counter-DDoS is possible practically instantly as soon as information about the C&C endpoints is available. However, the sustainability is negligible. The attacked botnet is disabled only as long as the counter-DDoS is executed. Also, the implications of launching an own DDoS attack need to be considered. It has to be ensured that legitimate services running in close proximity to the C&C endpoints are not adversely affected. In addition to that, DDoS attacks are illegal or even considered a hostile act in most countries.

2) Hack-Back

Another proactive countermeasure is hacking back, i.e. penetrating the C&C server and taking down the botnet from within. This requires the existence of a flaw in the C&C infrastructure which needs to be found and exploited. A team of highly skilled penetration specialists needs to be involved.
In open-source botnets, the C&C protocol can be easily discovered by analyzing the source code. Standard source code auditing tools can be used to find weaknesses in the code. Construction kit botnets are usually sold together with the C&C server, although it is typically in binary form. Therefore, reverse engineering

and binary code auditing skills are required. For specialized botnets it can be very difficult to obtain information about the C&C server. It is sometimes possible when using standard components with known vulnerabilities, e.g. specific Web servers. In all cases, analysts are required who are able to think outside box and identify non-obvious relationships between botnet components. Kit-based and specialized botnets require the highest reverse engineering skills.

The time required for such a hack-back differs among the different botnet classes. Because of the multitude of available code analysis tools, open-source botnets can often be hacked in a matter of minutes if a sufficient number of vulnerabilities exists, otherwise it is a matter of days depending on the complexity of the code. More time is required for kit-based botnets, since reverse engineering is needed most of the time. Because the server binary is available, offline and local stress tests can be performed. A minimum of several days can be expected, although the required time is more likely along the order of magnitude of weeks. Hacking of specialized botnets is very difficult. First the protocol has to be reverse engineered and possible weaknesses need to be derived. At least several weeks are needed for this.

Once access to the C&C server has been gained, diverse valuable information can be discovered. The installation of a root kit allows the complete control of the server machine and might even result in greater privileges than even the botherders have. However, in most countries it is illegal to gain access to computer systems of others without their knowledge. From an ethical point of view, hacking back is effectively fighting fire with fire.

3) Infiltration/Manipulation

Another proactive countermeasure is the infiltration of a botnet which might lead to the botnet being manipulated and/or disabled from within. This requires an in-depth understanding of the botnet's architecture and C&C protocol.

The skills needed vary for the different categories of botnets. Standard protocols, e. g. IRC and HTTP, can be automatically extracted, but especially for kit-based and specialized botnets, extensive reverse engineering skills are essential. Also, botnet domain knowledge coupled with out-of-the-box thinking is necessary to determine non-standard protocols. Cross-domain expertise is needed to identify and exploit weaknesses in the C&C protocol or architectures. Related fields in this case are communication protocol design, structured auditing as well as cryptography. Nevertheless, some manipulation vectors for standard protocols are well-known and can often be applied.

In terms of financial expenditure, analysis and monitoring environments need to be designed and built. Some organizations receive up to 100,000 malware samples per day. An investigation for the use of standard communication protocols takes place within a sandbox which has a minimum analysis time of 2 minutes. This requires around 140 machines running in parallel. Employing some heuristics allows the analysis to stop early. In addition to that, machines for monitoring are needed. Their number depends on the number of infiltrated botnets. Examples for existing frameworks for monitoring botnets are [17, 29].

The time required to infiltrate a botnet is difficult to estimate. A prerequisite is that malware samples are available for analysis. Their collection can already be a time-consuming task, especially if server-side polymorphism is used. Gaining an in-depth understanding of the botnet and its structures is also necessary. In case of standard protocols with standard manipulation vectors, a tactical takedown can be accomplished within minutes. The infiltration of botnets with non-standard protocols and the corresponding analyses can take up to several weeks, in lucky cases several days.
The sustainability of botnet infiltration is typically very high, provided it is not pursued too aggressively. For example, the aggressive monitoring of Storm by researchers was obvious to the botnet operators. If manipulations are made on the C&C server, they can be detected most of the time. To be truly effective, sudden strikes are essential.
The legal aspects of botnet infiltration still need to be investigated. From an ethical point of view, only the botnet's operation is interfered with. However, third-party data might be obtained or manipulated as a consequence, especially if the C&C is hosted on a hijacked system and depending on the architecture.

4) BGP Blackholing

Another possibility is the redirecting of botnet-related traffic, so-called sinkholing. The redirected traffic can simply be discarded or analyzed further to gather more information about infected machines. Resources with regard to money, skills and cross-domain knowledge are similar to those of regular C&C server takedowns. The processes can mostly be fully automated. However, the existence of backup channels for C&C processes can be challenging. Once sufficient information about the botnet and its structures is available, the C&C endpoints can be inserted into BGP feeds within a few seconds, although their propagation can take several minutes.

V. SUMMARY AND OUTLOOK

In this paper we have discussed the resources required for setting up and taking down botnets. In order to structure this we have categorized botnets into three groups: completely open-source botnets or those that use open-source components, construction kit-based botnets which are normally for sale, and specialized botnets tailored to a very specific task.
In general, kit-based botnets are the easiest to setup and operate since they were designed with user-friendliness in mind. When setting up an open-source botnet, basic software development skills are required which can be obtained in a matter of hours or days. Challenges can often be overcome by taking advantage of community support. This community support is missing for specialized botnets, often due to secrecy requirements.
Classical countermeasures are often inadequate when faced with intricate botnet structures and protocols. Proactive countermeasures are much better suited to deal with the botnet threat. Sufficient funds, time and development expertise in the area of malware analysis and reverse engineering are the most important requirements.

There is an increase in the amount of the respective required resources from open-source botnets through kit-based botnets to the specialized variants.
Currently, botnet operators are ahead in the arms race with security researchers, the anti-virus industry and law enforcement agencies. The currently performed anti-botnet activities are not as aggressive as they could be. This is partially due to lack of resources, the fear of legal consequences or uncoordinated efforts but also sometimes because of the fear of an intensifying arms race. Another reason is that monetary losses in the often targeted financial industry are still relatively moderate. However, studies show that there is a steady increase in the amounts lost due to credit card fraud, extortion and other botnet-related crimes. Thus, with a corresponding increase in funds for anti-botnet activities, it stands to reason that there will be more offensive botnet takedown attempts in the not-too-distant future, despite the fact that this would spark the feared arms race.

REFERENCES

[1] Symantec. Press release. Available online: http://www.symantec.com/about/news/release/article.jsp?prid=20090910_01, accessed February 2011.

[2] J. Nazario. *Politically Motivated Denial of Service Attacks*. In: C. Czosseck, K. Geers (Eds.), "The Virtual Battlefield: Perspectives on Cyber Warfare", IOS Press, 2009.

[3] Shadowserver. *60-Day Virus-Stats*. Available online: http://www.shadowserver.org/wiki/pmwiki.php/Stats/Virus60-DayStats, accessed February 2011.

[4] F. Leder, T. Werner, P. Martini. *Proactive Botnet Countermeasures – An Offensive Approach*. In: C. Czosseck, K. Geers (Eds.), "The Virtual Battlefield: Perspectives on Cyber Warfare", IOS Press, 2009.

[5] G. Klein, F. Leder, "Latest trends in botnet development and defense", Cooperative Cyber Defence Centre of Excellence, Tallinn, Estonia, 2010.

[6] *Metasploit - Penetration Testing Resources*. Available online: http://www.metasploit.com, accessed February 2011.

[7] *OpenSSL: The Open-Source Toolkit for SSL/TLS*. Available online: http://www.openssl.org, accessed February 2011.

[8] R. L. Rivest et al. *The MD6 hash function – A proposal to NIST for SHA-3*. Technical Report. Massachusetts Institute of Technology, Cambridge, MA, USA, April 2009.

[9] J. Gailly, M. Adler. *zlib Compression Library*. Available online: http://www.zlib.net, accessed November 2010.

[10] D. Fisher. *Storm, Nugache lead dangerous new botnet barrage*. Available online: http://searchsecurity.techtarget.com/news/article/0,289142,sid14_gci1286808,00.html, accessed December 2010.

[11] *VMProtect homepage*. Available online: http://vmpsoft.com, accessed November 2010.

[12] N. Villeneuve. *Tracking GhostNet: Investigating a Cyber Espionage Network*. Available online: http://www.scribd.com/doc/13731776/Tracking-GhostNet-Investigating-a-Cyber-Espionage-Network, accessed February 2011.

[13] N. Falliere, L. O. Murchu, E. Chien. *W32.Stuxnet Dossier*. November 2010. Available online: http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf, accessed February 2011.

[14] PC Plus. *Botnets Explained*. Available online: http://pcplus.techradar.com/feature/features/botnets-explained-30-09-10, accessed November 2010.

[15] Stevens, K., Jackson, D. *ZeuS Banking Trojan Report*. Available online: http://www.secureworks.com/research/threats/zeus, accessed December 2010.

[16] M86 Security Labs. *Web Exploits: There's an App for That*. Technical Report. Available online: http://www.m86security.com/documents/pdfs/security_labs/m86_web_exploits_report.pdf, accessed November 2010.

[17] Marco Cremonini and Marco Riccardi. *The Dorothy Project: An Open Botnet Analysis Framework for Automatic Tracking and Activity Visualization*. In *Proceedings* of the 2009 European Conference on Computer Network Defense (EC2ND '09). IEEE Computer Society, Washington, DC, USA, 52-54

[18] Kleissner, P. *AV Tracker homepage*. Available online: http://www.avtracker.info, accessed November 2010.

[19] Higgins, K. J. *Zeus Attackers Deploy Honeypot Against Researchers, Competitors*. Available online: http://www.darkreading.com/insider-threat/167801100/security/attacks-breaches/228200070/index.html, accessed December 2010.

[20] Websense, Inc. *Websense 2010 Threat Report*. Technical Report. Available online: http://www.websense.com/content/threat-report-2010-introduction.aspx, accessed November 2010.

[21] Damballa. *Want to rent an 80-120k DDoS Botnet?*. Available online: http://blog.damballa.com/?p=330, accessed February 2011.

[22] M. Bailey, E. Cooke, F. Jahanian, Y. Xu, and M. Karir. *A survey of botnet technology and defenses*. In Proceedings of the 2009 Cybersecurity Applications & Technology Conference for Homeland Security, pages 299–304, Washington, DC, USA, 2009. IEEE Computer Society.

[23] P. Bächer, T. Holz, M. Kötter, and G. Wicherski. *Know your enemy: Tracking botnets*. Honeynet Project KYE series, 2007.

[24] F. Leder and T.Werner. *Don't do this at home - owning botnets*. In T2 information security conference, Helsinki, Finland, 2009.

[25] McAfee, Whitepaper, *Global Energy Cyberattacks: "Night Dragon"*, Version 1.4, February 10, 2011, http://www.mcafee.com/us/resources/white-papers/wp-global-energy-cyberattacks-night-dragon.pdf , accessed February 2011

[26] D. Fisher. *Storm, nugache lead dangerous new botnet barrage*. http://searchsecurity.techtarget.com/news /article/0,289142,sid14_gci1286808,00.html, , accessed February 2011

[27] Spencer Kelly, BBC, *Gaining access to a hacker's world*, 13 March, 2009, http://news.bbc.co.uk/2/hi/programmes /click_online/7938201.stm , accessed February 2011

[28] Abuse.ch, *Zeus Tracker*, https://zeustracker.abuse.ch/ , accessed February 2011

[29] G. Wicherski, *botsnoopd - Efficiently Spying on Botnets*, GovCert Symposium, September 16, 2008, Rotterdam, NL

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

Preserving Organizational Privacy in Intrusion Detection Log Sharing

Hayretdin Bahşi
Turkish National Research Institute of Electronics and Cryptology
Kocaeli, Turkey
e-mail: bahsi@uekae.tubitak.gov.tr

Albert Levi
Faculty of Engineering and Natural Sciences
Sabancı University
İstanbul, Turkey
e-mail:levi@sabanciuniv.edu

Abstract- **This paper presents a privacy-preserving framework for organizations that need to share their logs of intrusion detection systems with a centralized intrusion log management center. This centralized center may be an outsourced company that provides an intrusion detection management service to organizations or a system of the National Computer Emergency Response Team that probes the attacks targeting organizations that have critical information systems. For reasons of ensuring privacy, we adopt the notion of *l*-Diversity in the course of collecting intrusion logs from organizations. Within our framework, an organization ensures the people in the center cannot discern the exact origin of any intrusion log among the other *l*-1 organizations. Also, it is not possible to precisely identify the classification type of an intrusion log from among other *l*-1 types. Within this framework, the intrusion log management center can analyze the anonymous data, since the proposed privacy preserving solution creates little information loss. If required, it sends an alarm to the appropriate organization within a reasonable time. The center has the option of publishing useful information security statistics about specific organizations or about the whole ecosystem by using the privacy preserved intrusion logs.**

Keywords: privacy preserving framework, intrusion detection, log sharing

I. INTRODUCTION

It is known that hackers share information with each other in order to attack victims. In underground communities, zero day vulnerability information, target victim information, stolen credit card numbers, bots, spam mail lists, attack tools, etc. are shared or sold easily. On the other hand, system managers, who strive to defend their systems against hackers, need to share related materials about defensive tools, methods and information. The defensive experience of an organization can easily be transferred to others by sharing intrusion detection system logs.

It is common that most of the organizations somehow use intrusion detection systems to detect attacks against their systems. These systems do not always produce useful outputs. In particular, the elimination of false positive alarms

requires labor intensive work. An information security expert has to choose the set of attack signatures that are appropriate for his system and eliminate false positive alarms. However, most organizations cannot reserve staff for this task due to a lack of specialized technical personnel or due to a lack of budgetary funding. Under these circumstances, the outsourcing of intrusion log analysis could be a good alternative.

Nowadays, National Computer Emergency Response Teams (NCERT) are determining ways to perform proactive nationwide security countermeasures in order to detect and prevent cyber attacks targeted at national critical information infrastructures. These infrastructures generally belong to different organizations. NCERTs try to determine ways to centrally probe them. Probing aims to deduce the overall threat state of each organization and determine the overall threat level of the country. For this aim, a distributed intrusion detection system has to be setup and managed. Moreover, collected intrusion logs have to be centrally stored and analyzed.

In the above both cases, there is a need for a central intrusion log management office (CILMO) to store logs of different organizations centrally, analyze them, detect attacks, send alarms to organizations and generates statistics for determining nationwide threat levels.

The primary obstacle in forming a CILMO is the privacy concerns of organizations. Intrusion logs contain valuable information about organizations, such as detailed knowledge of targeted information assets, attack times, types of attacks, results of attacks, etc. Organizations are reluctant to share intrusion logs due to two main reasons. First, they do not fully trust the personnel of CILMO, because administrators of CILMO may intentionally misuse their attack information. The second reason may be the lack of appropriate security and privacy countermeasures, which have to be applied to the intrusion logs during their transmission, processing and storage. Without solving these security and privacy problems, organizations generally do not wish to send their intrusion logs to a CILMO, even though it may have been set up by a NCERT team.

Organizations are confronted with the dilemma between privacy risks and the benefits of sharing intrusion logs. Therefore, one has to deal with the trade-off between privacy and information loss, according to the needs of organizations.

In this paper, a privacy-preserving framework based on *l*-diversity is presented for intrusion log sharing. This notion guarantees that the exact classification type of an intrusion log cannot be identified among other *l*-1 types. Also, privacy schema enables us to hide the source organization of the intrusion log among *l*-1 organizations. Through the collection of privacy-preserved intrusion logs, this framework enables CILMOs to perform detailed security analysis of organizations, draw conclusions about the general security status of organization categories and prepare a warning mechanism.

The general structure of the paper is as follows: Section II gives some background information and introduces the threat and network model. Section III details the proposed anonymization method. Section IV gives the results of experiments performed in evaluation of the proposed method. Section V concludes the paper.

II. MOTIVATION AND BACKGROUND

A. *k-Anonymity and l-diversity*

Privacy problem cannot be easily solved by merely removing identity information (name, social security number, etc.) from the records of individuals. Data fields called quasi-identifiers may be used to identify a person by using external information sources. This attack technique is called “Re-identification attack” [1] or “record linkage” [2]. For example, in a hospital database, address, sex or other attributes can precisely identify an individual. *k*-Anonymity [1], which is defined as being not identifiable of an individual within a set of *k*-1 individuals, is used as a privacy criterion in order to make data resistant to re-identification attacks. *k*-Anonymity generalizes or suppresses quasi-identifiers of data records so that an individual cannot be differentiated between other records of *k*-1 individuals by using those quasi-identifiers.

It has been shown that without finding the exact owner of a record, if sensitive attribute exists in a record, it may be possible to identify the sensitive attribute of an individual in some circumstances by an attack called an “attribute linkage attack" [2]. Sensitive attribute includes information such as the health of a patient in a hospital database. In order to prevent this problem, *k*-anonymity notion extended in some studies. Machanavajjhala et. al. extended *k*-anonymity with a *l*-diversity notion in order to cover these attacks [3]. In addition to *l*-diversity notion, *p*-sensitivity and *t*-closeness notions are proposed [4], [5].

B. *Threat and Network Model*

In our study, organizations send their intrusion logs to a trusted party. In a realistic scenario, a trusted party may be an Internet service provider (ISP). Normally, all the network traffic between the Internet and organizations is managed by ISPs. Organizations legally protect themselves against the possible malicious activities of ISP administrators by service level agreements, which include non-disclosure and security protection terms. ISPs can be presumed to be trusted parties due to these agreements.

A sample system topology for the proposed privacy framework is given in Figure 1. A trusted party anonymizes intrusion logs, strips off the destination IP information of a log and appends a destination tag instead of the destination IP, which only represents the source organization. Target Service, source IP and detection time attributes are classified as quasi-identifiers and intrusion classification is accepted as sensitive attribute. According to this attribute classification, our anonymization method provides the prevention of record and attribute disclosure by providing *l*-diversity property of intrusion logs.

It is assumed that in each log originating from organizations, pre-exploitation, exploitation and post-exploitation activities are correlated and one log entry is created for each attack. If one attack targets the many servers of an organization, only one log entry is produced by IDS.

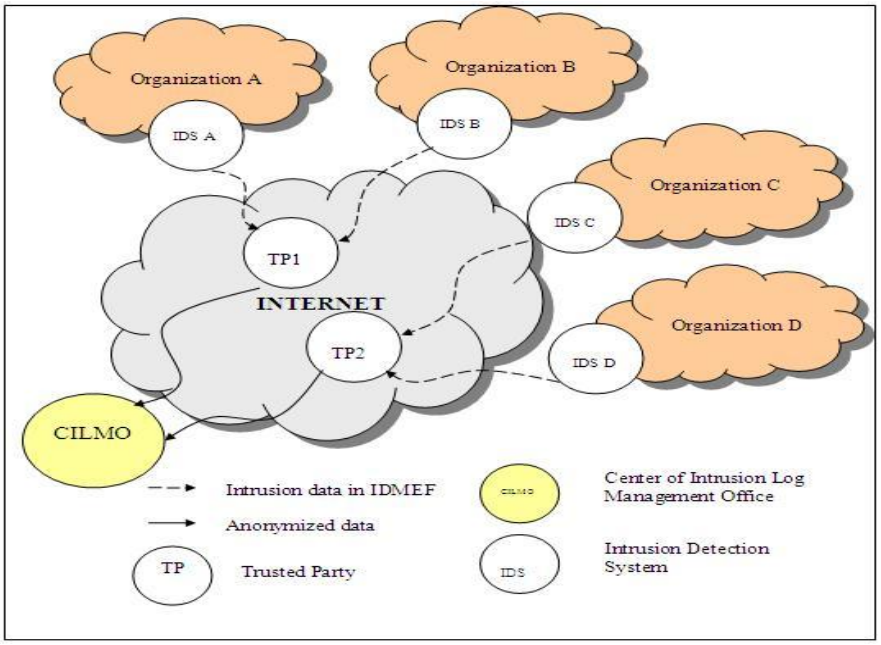

Figure 1. Figure 1. System Topology for Privacy Framework

C. *Related Work*

Some organizations implement intrusion log collection systems for determining the general security level of the Internet. Deepsight Threat Management System [6], which is managed by Symantec, gives information to its customers about emerging threats, vulnerabilities, risks, etc. The system does not use any anonymization method during data collection. Internet storm center, which is implemented by SANS [7], collects intrusion detection system and firewall logs from volunteer organizations producing general analysis results for the public, and creates customized warning information for organizations. They just simply remove the identifying parts of intrusion data by masking the destination IP of logs.

There are studies about anonymzing the IP address of network logs. In a basic solution, actual IP addresses are replaced by a randomly selected IP addresses according to a permutation function. New random IP addresses do not even contain the sub-net information.

Truncation is another anonymization method that converts a fixed number of the least significant bits of an IP address to zero. This means that the remaining information can show only the subnet or network class information of IP addresses. From anonymized data, anyone can deduce the subnet information but cannot determine whether logs belong to a particular subnet.

In *prefix-preserving pseudonymization*, which is adapted in *TCPdriv* [8], IP addresses are mapped to pseudorandom anonymized IP addresses by an

anonymization function that uses common tables. Reference [9] proposed a prefix-preserving pseudonymization method, Crypto-PAn, which works consistently in multiple traces by using a shared key. Crypto-PAn is re-implemented in Java for the anonymization of Netflow logs [10]. The anonymization of all fields of Netflow and syslog data for sharing them with managed security service providers is performed in [11].

III. PROPOSED ANONYMIZATION METHOD L-ACM

k-ACM (*k*-Anonymous Clustering Method) is proposed in [12], which k-anonymizes the data by using the hierarchical bottom-up clustering method. This method is applied for the anonymization of collected data in wireless sensor networks [12], [13]. In this paper, *k*-ACM is modified for the anonymization of intrusion logs in order to make them *l*-diversity. The proposed method is referred to as the *l*-diversity Anonymous Clustering Method (*l*-ACM).

Subsection III.A explains how the collected information is represented in our proposed method. In Subsection III.B, distance metric, which is used in the clustering process, is described. Subsection III.C presents details on the bottom-up clustering process, which is the core of the proposed method.

A. Data Representation

l-ACM uses the data representation model used in [12], [13]. This subsection describes the details of this model. Suppose input data is a table T with m attributes, r records. Tij represents the j'th attribute of the i'th record where $\{i : 1 \leq i \leq r\}$ and $\{j : 1 \leq j \leq m\}$. Table T is represented by a set of bit strings B, where Bij is a bit string representation of j'th attribute of i'th record. The k'th bit of Bij is shown as $Bij(k)$.

Suppose that the j'th attribute of a table is categorical and there are dj distinct values. These values are indexed by k and shown as $Vj(k)$ where $\{ k : 1 \leq k \leq dj \}$. The bit string of this categorical attribute has a size of dj and is formed as follows:

$$\text{If } Tij = Vj(k) \text{ then } Bij(k) = 1 \text{ else } Bij(k) = 0 \text{ as } \forall k : 0 \leq k \leq dj ,$$

If the attribute is numerical, the range of the attribute is divided into equal-sized intervals and each interval constitutes a categoric value.

B. Information Loss Metric

In order to evaluate the quality of data, *l*-ACM uses the information loss metric of *k*-ACM [12]. This metric basically uses the entropy concept of the information theory [14]. Information loss is quantified by the difference of entropies between the *l*-diversified data and the original data.

Assume that input data, T, has r records and m attributes. B is the bit string representation of data set, T. C is the random variable that gets the probability value of an attribute value in a *l*-diversified data entry, which is the actual attribute value in the original data. B is normalized by the number of bits that have

value “1” (from here on we use “true bit” to refer to a bit that has the value “1”). Normalized version forms data set $\overline{B}$.
Information loss of a data table *T*, *IL*(*T*), is equal to the conditional entropy, $H(C \mid B)$. Here, the conditional entropy gives the uncertainty about the prediction of the original attribute values of a record when we have the knowledge of corresponding *l*-diversified bit strings of that record. The original data has only one true bit. Anonymization increases the number of true bits. Each true bit actually represents the possible original attribute value. As the number of true bits increases, disorder of the data increases because it is harder to predict which one of them is the original true bit. Conditional entropy $H(C \mid B)$, which is equal to the information loss of table *T*, *IL*(*T*), can be determined as follows:

$$\begin{aligned} IL(T) = H(C \mid B) &= \sum_{B_{ij} \in B} p(B_{ij}) H(C \mid B = B_{ij}) \\ &= - \sum_{B_{ij} \in B} p(B_{ij}) \sum_{k \in \{1..z\}} p(C = k \mid B_{ij}) \log p(C = k \mid B_{ij}) \end{aligned} \quad (1)$$

In Eqs. (1), it is assumed that each attribute is converted to bit strings of the size *z*. This means that all categorical attributes have *z* distinct attribute values and all numerical attributes have *z* number of interval ranges. Also, it is assumed that all *k*’s, where the equalities of $p(C=k|Bij)=0$ are true, are excluded from the summation. *C* random variable can take values from the set $\{1..z\}$. Actually, $\overline{B}$ is calculated for determining the value of this random variable.

$$p(C = k \mid B = B_{ij}) = \overline{B}_{ij}^{k} \; for \; each \; k : 1 \le k \le z \quad (2)$$

In Eqs. (2), it is assumed that each record has equal probability to be chosen and each attribute of record has the same probability. Therefore, the probability mass function of the *j*’th attribute of the *i*’th record, $p(Bij)$ is calculated as $p(Bij)=1/m.r$. Eqs (1) can be rewritten as follows:

$$IL(T) = - \sum_{B_{ij} \in B} \frac{1}{m.r} \sum_{k \in \{1..z\}} \overline{B}_{ij}^{k} . \log(\overline{B}_{ij}^{k}) \quad (3)$$

Suppose that *F* is the array that contains the number of true bits of the bit string array *B*. The total number of true bits in Bij is Fij. The total number of elements in $\overline{B}ij(k)$ that have the value of $1/F_{ji}$ is equal to F_{ji}, and the rest are zero. Therefore, the second sum operation of Eqs. (3) yields the value, $logl/F_{ji}$. The

simplest equation for the information loss of data table T, $IL(T)$, can be calculated as follows:

$$IL(T) = -\sum_{F_{ij} \in F} \frac{1}{m.r} \log\left(1/F_{ij}\right) == \frac{1}{m.r} \sum_{F_{ij} \in F} \log(F_{ij}) \quad (4)$$

C. *Bottom-up Hierarchical Clustering Process*

Method bases on forming clusters of input vectors iteratively. Each cluster numerated as C_j^l in each epoch, l, contains a number of input vectors, N_j^l, and a representative vector, R_j^l where j is the index number of cluster. Suppose that the k^{th} data item of the representative vector is denoted as $R_j^l[k]$. The representative vector is actually the anonymized output of input vectors belonging to the cluster that is formed by generalization operations of some data parts of vectors.

The hierarchical clustering process begins with the assumption that each input vector constitutes a separate cluster and that vector is also a representative vector of the cluster. In each epoch, by using the information loss metric described in Section III.B, distances between each cluster are calculated. The distance between any two clusters is actually equal to the information loss that may occur if both clusters are merged.

The two clusters that have the smallest distance, e.g. clusters C_s^l and C_t^l, are chosen for merging. The new bigger cluster, C_u^{l+1} which contains the vector items of both clusters, is formed and the former two clusters are deleted. N_u^{l+1} is equal to the sum of N_s^l and N_t^l . Anonymization is performed by generalization. $R_u^{l+1}[k]$ is equal to the XOR of $R_s^l[k]$ and $R_t^l[k]$.

l-ACM keeps on clustering iterations up to the point where each cluster contains a record set that has distinct l sensitive attribute values and l different sets of quasi-identifier attributes. Representative vectors of remaining clusters form the l diversified outputs.

A target organization can be considered as an identifier of an intrusion log. In our case, CILMO needs the names of the target organizations in order to perform the required security analysis tasks. The names of the target organizations are transferred to CILMO in such a way so that nobody can deduce the name of the exact organization of an intrusion log among the l-1 organizations.

The same organization may send many intrusion logs to CILMO. If one anonymity set produced by l-ACM has many intrusion logs of the same organization, this situation may violate l-diversity property. Therefore, l-ACM guarantees that each

record in each cluster has to belong to a different organization. Suppose that n is the number of records. The set of all target organizations is represented as $\{O_1, O_2, ..., O_n\}$. Assume that all records have m different sensitive attribute values where $m > l$ and these attributes values are $\{ S_1, S_2, ..., S_m \}$. The data sent to CILMO can be shown as $\{O_1, O_2, ..., O_n\}, R_i, \{ S_1, S_2, ..., S_m \}$.

A running example of l-ACM is shown in TABLE I and TABLE II. Assume that each destination IP belongs to a different organization. The destination IP of the intrusion log is replaced with the name of the organization during anonymization. The trusted party gathers the original data shown in TABLE I, produces three clusters that each have two elements and makes the data 2-diversified. Each row in this table represents one cluster. All the attributes are converted to sets of distinct attribute values. l-ACM guarantees that in the destination organization attribute, two distinct organization names exist and the classification attribute consists of a set that has two different classification values. Since the source IP, time and destination port attributes are chosen quasi-identifiers, l-ACM tries to minimize the number of distinct attribute values of these attributes in anonymized output.

TABLE I. AN EXAMPLE OF THE ANONYMIZATION OF INTRUSION LOGS – ORIGINAL DATA

Dst IP	Src IP	Time	Dst Srv	Classification
201.2.1.10	195.100.4.4	11:00	53	DNS Zone Transfer
223.23.5.4	195.100.4.4	11:30	8080	WEB IIS ISAPI
212.125.12.12	198.166.3.3	11:40	3372	DoS MSDTC
222.19.1.103	190.67.30.3	11:45	1543	NETBIOS SMB
208.234.3.105	199.201.45.56	11:55	80	WEB-COLDFUSION
200.188.5.17	191.34.32.1	12:05	1548	DOS IGMP

TABLE II. AN EXAMPLE OF THE ANONYMIZATION OF INTRUSION LOGS – 2-DIVERSIFIED DATA

Dst IP	Src IP	Time	Dst Srv	Classification
{ O1, O2}	{195.100.4.4}	{11:00, 11:30}	{8080, 53}	{DNS Zone Transfer, WEB IIS ISAPI}
{ O3, O4}	{198.166.3.3, 190.67.30.3}	{11:40, 11:45}	{1543, 3372}	{DoS MSDTC, NETBIOS SMB}
{ O5, O6}	{199.201.45.56, 191.34.32.1}	{11:55, 12:05}	{80, 1548}	{WEB-COLDFUSION, DOS IGMP}

D. *Warning Mechanism*

CILMO may need to warn organizations about a very critical intrusion. Likewise, if the proposed anonymization method is used in intrusion log sharing, CILMO does not know the exact intrusion classification for the exact originator. It only knows that a set of organization corresponds to a set of intrusion classification values. CILMO may be interested in one intrusion classification among these values. If it is assumed that the trusted party does not store any information including the mappings of original data with anonymous data, the warning can be performed by only distributing it to each IDS management server of all candidate

organizations. The details of the warning mechanism are described through an example in Figure 2. Each organization sends their logs, which are labelled as r_1, $r_2, .., r_6$ to the trusted party (TP), in step 1. TP anonymizes the data according to 2-diversity criteria and sends the anonymous outputs a_1, a_2, a_3 to CILMO in step 2. Assume that CILMO decided to warn the organizations about the DNS Zone Transfer attack due to its seriousness. Assume that r_1 has this classification type. CILMO chooses the anonymous record (a_1) which has this attack type from the set of classification attributes. CILMO creates w_1 from a_1 by stripping off all organization attributes and all classification information except "DNS Zone Transfer" and sends w_1 to IDS management servers of organization 1 (O1) and organization 2 (O2) in steps 3 and 4. In step 5, O1 and O2 query whether an intrusion log exists about the profile given in w_1 and determine whether the corresponding warning is related with their organization.

A drawback of this mechanism is that the organization O2, which decides the warning, does not belong to itself in the above example. It also receives the profile information of the intrusion that occurred for O1 without knowing the targeted organization is O1.

If the trusted party is allowed to store mapping information between original data and anonymous output, after deciding the warning message, CILMO sends w_1 to TP. TP finds the exact intrusion log record that matches with w_1, deduces that it is r_1 and relays r_1 to O1. In this method, an organization does not learn anything about the intrusion logs of other organizations. The warning is sent directly to the owner organization of the intrusion log.

IV. PERFORMANCE EVALUATION OF L-ACM

In this part, the performance of *l*-ACM is evaluated in terms of *information loss* and the *average response time* of intrusion log records. The average response time, T_{avg}, shows the average amount of times between the generation of the log at the owner organization and the arrival of the corresponding warning to that organization from CILMO.

In our experiments, each organization generates an intrusion log in a such a way that all the attributes of logs are formed using uniform distribution. The log generation time for i^{th} log record is represented as t_g^i. Log generation rate, *lgr*, which is the number of produced logs per minute, is a predetermined parameter that adjusts the speed of log generation. It is assumed that each organization uses the same log generation rate. All log records generated in one minute are collected at the organization site and they are sent to CILMO at the end of that minute. Therefore, i^{th} log record waits 60 - t_g^i seconds at the organization site before being sent to CILMO. After CILMO receives the logs, the anonymization operations take place by using *l*-ACM. Anonymization is completed in several steps. In each step, the data set that includes only one record from each organization is chosen among the received logs and they are anonymized.

Otherwise, if we include more than one record from each organization, an anonymity set may contain more than one record belonging to same organization, which violates *l*-diversity property. The restriction of one record from the same organization actually means that the number of steps needed for completion of anonymization is numerically equal to the log generation rate. The duration of the m^{th} anonymization step is represented as t_a^m.

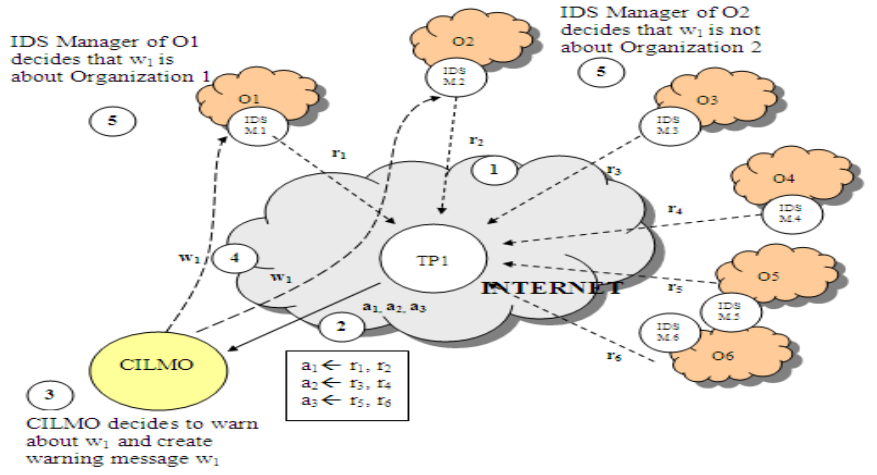

Figure 2. Figure 2. Warning Mechanism with the requirement that trusted party does not store any information

In *l*-ACM, we use a record selection method for preparing the input data of each anonymization step. Our method chooses an initial record from the first organization. For each other organization, the logs of an organization are compared with the record of the first organization and the one that bears the most similarity is chosen as an input record in that step.

Anonymized outputs are analyzed by CILMO. If analysis results require the sending of a warning to the appropriate organization, warnings are sent by using one of the methods given in Section III.D. In performance calculations, a parameter called log analysis time, t_l, is used for the log analysis of one log record at CILMO. Warnings are sent after this analysis time has passed.

The transmission time needed for transferring one log record from the organization to CILMO and the time for transferring one warning to the organization is represented as t_r. In average response time calculations, we assume that for each log record, CILMO sends a warning message. The average response time for a log record is calculated as given in Eqs. (5). We assume that the total number of the input record is *n*.

$$T_{avg} = (60 - t_g^i) + \sum_{s=1}^{s=m} t_a^m + t_l + 2.t_r \quad (5)$$

The effects of changes in parameter *l* and *lgr* with respect to information loss and response time performances of *l*-ACM, are investigated via simulations. Experiments are performed in a laptop that has 1.20 GHz CPU and 2GB RAM. Intrusion data is synthetically generated. A java implementation is developed for data generation, the application of *l*-ACM and evaluating the results.

k-ACM calculates the information loss according to Eqs. (4). In this formula, F_{ij} is the total number of bits that have the value of '1' for the *i*th record of *j*th attribute. On the other side, *l*-ACM produces anonymized output with an attribute value sets instead of bit strings. Therefore, *l*-ACM uses the size of the attribute value set (which means the number of distinct elements in the set) instead of F_{ij} .

There are 100 distinct attackers in the network. The number of distinct values for intrusion classification is 15 and the number of slots for time value is 100. There are 10 distinct destination services in the data set. According to these parameters, maximum information loss is calculated as 5.54 via the help of Equation 4.

The effects of *lgr* and *l* values on information loss results is given in Figure 3. In these experiments, the number of organizations that send their logs to CILMO is fixed to 500. As shown in Figure 3, increase in *lgr* does not affect information loss values for each *l* value. The effects of *lgr* and *l* values on average response time are given in Figure 4. In this experiment, number of organizations is also fixed to 500. It is observed that the average response time increases as *lgr* increases for each *l* values. There is a linear relationship between the average response time and *lgr* values. Since *lgr* also determines the number of anonymization steps performed at CILMO, an increase in the number of steps increases the time for anonymization operations. For the same *lgr*, we get higher than average response time values for higher *l* values due to the need for much more processing in hierarchical clusterings.

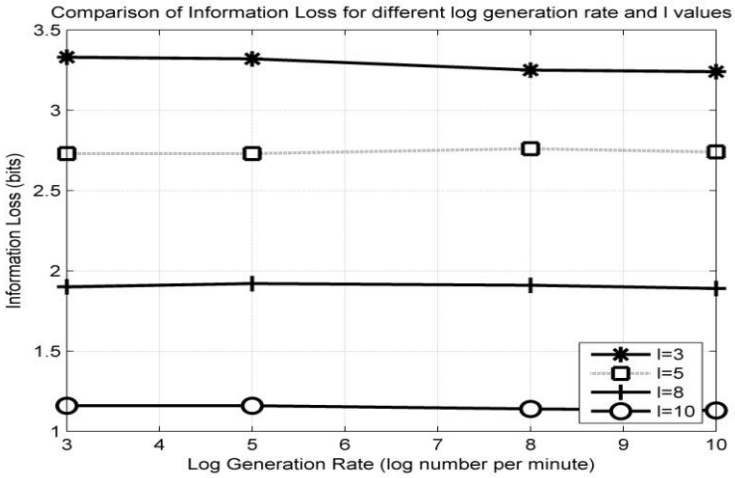

Figure 3. Figure 3. Effects of *lgr* and *l* on Information Loss

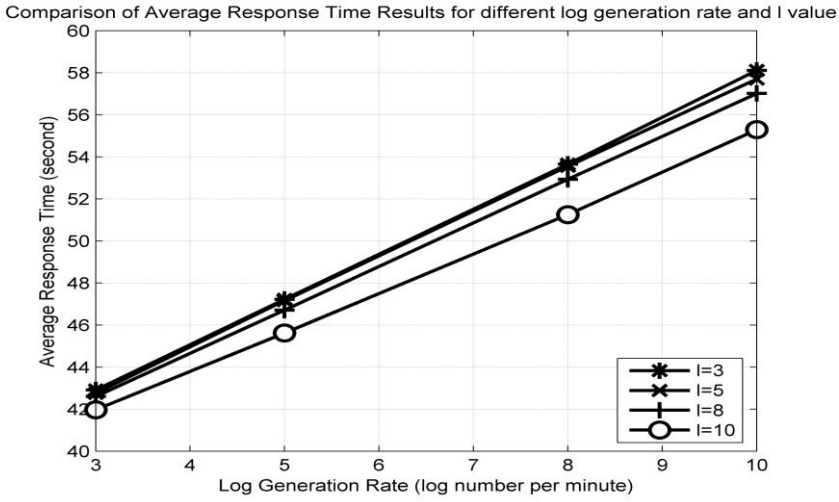

Figure 4. Figure 4. Effects of *lgr* and *l* on Average Response Time

Effects of the changes in the number of organizations are analyzed. Figure 5 shows the effects of organization number to information loss. Figure 6 analyzes the average response time results of *l*-ACM with a different number of organizations. In these experiments, the *l* and *lgr* values are fixed to 5 and 8 respectively.

From Figure 5, it is deduced that the information loss value decreases as the number of organizations increases. Since, anonymization is performed among bigger sets of log records in higher organization numbers; *l*-ACM has the possibility to find more similar records during hierarchical clustering. However, the decrease is very small according to experimental results.

Figure 6 shows that a higher number of organizations cause higher response times. There exists an exponential increase in response times. An increase in the number of organizations means higher number records are given as an input to *l*-ACM in each anonymization step.

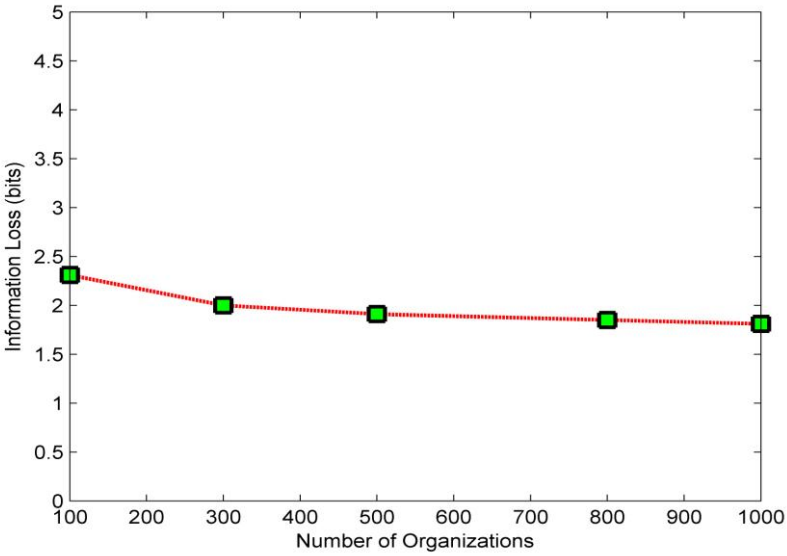

Figure 5. Figure 5. Effects of Organization Number on Information Loss

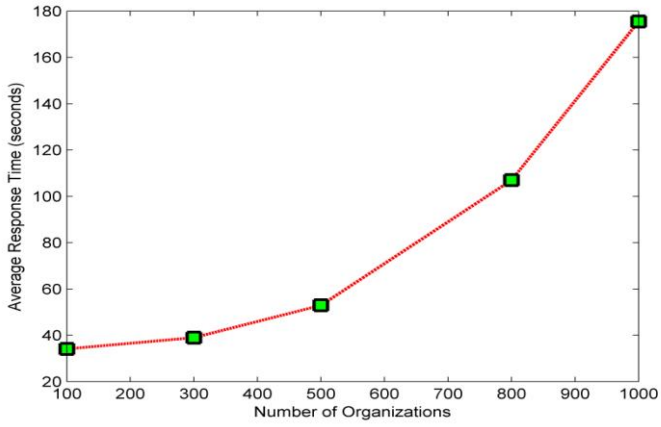

Figure 6. Figure 6. Effects of Organization Number on Average Response Time

V. CONCLUSION

In this paper, the privacy preserving framework is proposed for the collection of intrusion logs from different organizations through a central intrusion log management office. This office is tasked for determining the overall security posture of the whole organization ecosystem, the designation of the security status of monitored organizations, and it gives feedback or warnings to organizations about critical intrusions. The privacy threat model states that the collected log has to have *l*-diversity property. This means, any administrator of the central office cannot deduce the exact classification type of intrusion log among the *l* classification types. *l*-ACM (*l*-Diversity Anonymous Clustering Method), is proposed for this purpose. Different warning mechanisms are presented according to the security requirement on whether trusted parties are allowed to temporarily store network traffic.

REFERENCES

[1] L. Sweeney, "k-anonymity: A model for protecting privacy," Int'l Journal on Uncertainty, Fuziness, and Knowledge-based Systems 10(5),

[2] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: A survey on recent developments," *ACM Computing Surveys* , 2009.

[3] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," in Proceedings of 22nd International. Conference on Data Engineering, p. 24, ICDE, 2006.

[4] M. T. Truta and V. Bindu, "Privacy protection: *p*-sensitive *k*-anonymity property," in *Proceedings of the Workshop on Privacy Data Management*, p. 94, Workshop on Privacy Data Management, In Conjunction with 22th IEEE International Conference of Data Engineering (ICDE), (Atlanta, Georgia), 2006.

[5] N. Li, T. Li, and S. Venkatasubramanian, "*t*-closeness: Privacy beyond *k*-anonymity and *l*-diversity," CERIAS Tech. Report 2007-78, Purdue University, 2007.

[6] "Deepsight threat management system." https://tms.symantec.com/Default.aspx.

[7] "Internet storm center." http://isc.sans.org/.

[8] G. Minshall, "Tcpdriv command manual," 1996.

[9] J. Xu, J. Fan, M. H. Ammar, and S. B. Moon, "Prefix-preserving ip address anonymization: Measurment-based security evaluation and a new crypthography-based scheme," IEEE International Conference on Network Protocols, 2002.

[10] A. Slagell, Y. Li, and K. Luo, "Sharing network logs for computer forensics: A new tool for the anonymization of netflow records," Computer Network Forensics Research Workshop, held in conjuction with IEEE SecureComm, 2005.

[11] J. Zhang, N. Borisov, and W. Yurcik, "Outsourcing security analysis with anonymized logs," 2nd IEEE Intl. Workshop on the Value of Security through Collab., 2006.

[12] H. Bahsi and A. Levi, "k-anonymity based framework for privacy preserving data collection in wireless sensor networks," *Turkish Journal of Electrical Engineering and Computer Science* 18(2), pp. 241–271, 2010.

[13] H. Bahsi and A. Levi, "Data collection framework for energy efficient privacy preservation in wireless sensor networks having many-to-many structures," *Sensors* 10(9), pp. 8375–8397, 2010.

[14] P. Andritsos and V. Tzerpos, "Software clustering based on information loss minimization," in *Proceedings of 10th Working Conference on Reverse Engineering*, p. 334, WCRE 03, 2003.

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

Requirements for a Future EWS – Cyber Defence in the Internet of the Future

Mario Golling and Björn Stelte
Universität der Bundeswehr
Faculty of Computer Science
D-85577 Neubiberg, Germany
Email: {mario.golling and bjoern.stelte}@unibw.de

***Abstract*- The emergence of new technologies and services as well as trillions of devices and petabytes of data to be processed and transferred in the Internet of the Future mean that we have to deal with new threats and vulnerabilities, in addition to handle the remaining old ones. Together with the rise of Cyber Warfare and the resulting impact on the environment means that we have to bring intelligence back to the network. Consequently, effective Cyber Defence will be more and more important. In this paper we will show that the proposed requirements for an Early Warning System are a main part of future Cyber Defence. Special attention is given on the challenges associated to the generation of early warning systems for future attacks on the Internet of the Future. The term Cyber War is used frequently but unfortunately with different intends. Therefore, we start with a definition of the term Cyber War focusing on security aspects related to the Internet of the Future, followed by an exemplification of a Cyber War, of its implications and the challenges associated to it. Then we proceed with an analysis of state of the art recent work that has been proposed on the topic. Additionally the weaknesses of these analyzed systems and approaches are presented. Finally we propose guidelines and requirements for future work which will be needed to implement a next generation early warning system for securing the Internet of the Future.**

Keywords: Cyber Defence, Cyber Warfare, Internet of the Future, Early Warning System

I. INTRODUCTION

Although it is not exactly known how the Internet of the Future will look like, some challenges are quite obvious. Cloud computing allows that data and services are provided somewhere in the network; the Internet of Things indicates an enormous amount of devices to be managed as well as data to be processed; privacy requires that a high amount of packets (payload) needs to be encrypted; Security management demands to develop concepts to assure trustworthiness and manage the security capabilities consistently according agreed security policies in the future.

Not only the Internet will change significantly, also attacks on the Internet will change dramatically. In the last years Cyber Attacks became more and more visible [1]. It's generally acknowledged that the amount of these Cyber Attacks will continue to increase. Beside of the number of attacks also the impact of future Cyber Attacks are more harmful. Over the past few years more and more mission critical devices like critical infrastructures [2] are accessible within the Internet and the people behind the attacks are more skilled then before. Of these people a huge amount is professional trained on Cyber Warfare, like the U.S. Cyber Command. It's not surprising that many nations establish cyber commands because future acts of war will also have an impact on the Internet. Warfare is shifting more and more from the traditional battlefield into the digital battlefield. Consequently, Cyber Defence is essential for every nation. Due to the distributed nature of the Internet a Cyber Attack will almost never attack only one nation. Therefore, a cooperative defence strategy is needed to thwart the impact of the attack.
Traditionally Internet providers are using Early Warning Systems (EWS) to protect themselves against and quickly react on certain Cyber Attacks. Due to a new level of cyber threats we need an improved EWS architecture with the requirement not to be limited to the borders of different providers, based on traditional packet inspection, but to gather, analyze and correlate available (network) data (e.g. flows) to detect, analyze and react to threat patterns in near real time. This includes the development of completely new approaches such as the development of virtual sensors, sophisticated correlation of data, new reasoning models for network behavior analysis, learning algorithms as well as concepts to deal with scalability, dependability and resilience, especially in IPv6 networks.

II. DEFINITION OF TERMS

Before going deeper into descriptions about principles and patterns of Cyber Wars, we define the terms used within this publication. In comparison to kinetic warfare, which we define as warfare practiced in the "real world" by all the tanks, ships, planes and soldiers of current militarizes, we like to define Cyber Warfare based on the two definitions of John Arquilla/David Ronfeldt [3] and Richard A. Clarke/Robert Knake [4] as follows:
Cyber Warfare is the unauthorized conducting of a penetration - including the preparation - by, on behalf of, or in support of, a government into another nations's computer or network, or any other activity affecting a computer system, in which the purpose is to add, alter, falsify or delete data, or cause the disruption of or damage to a computer or network, or the objects a computer system controls (such as SCADA-systems "supervisory control and data acquisition").
In contrast to classical Cyber Attacks, where the intentions are almost similarly, the implications of a Cyber War are highly relevant, because attacking a nation - and in consequence attacking the critical infrastructure of a nation - creates a higher impact and is most likely not limited to the borderline of that nation.
Following the ideas of Clausewitz on war [5] (who saw the war primarily as a clash between nations) the sophisticated threat level of Cyber War is not individuals or companies, it is one or more nations.

III. DEFINITION OF TERMS

A. *Motivation*

For a better understanding of the concepts of Cyber Wars and the connection to our topic, we like to quickly illustrate the Cyber Assault against Estonia in 2007 [6], [7], [8]. During the night of 26 April to 27 April 2007, now known as Bronze Night, riots broke out in the Estonian capital Tallinn after the Estonian government moved the Bronze Soldier - a memorial statue honoring Soviet World War II war dead - from the central square of Tallinn, to a cemetery on the city's outskirts. The move also ignited nationalist responses in the Moscow media. Simultaneously, the conflict moved into cyberspace.

The attacks began on April 27. On Russian language Internet forums, Estonian officials say, instructions were posted on how to disable government Web sites by overwhelming them with traffic, a tactic known as a denial of service attack. Most of the attacks were Internet Control Message Protocol (ICMP) floods lasting 10 hours or more. The Web sites of the Estonian president, the prime minister, Parliament and government ministries were quickly swamped with traffic, shutting them down. Hackers defaced other sites, putting, for instance, a Hitler mustache on the picture of Prime Minister Andrus Ansip on his political party's Web site. The assault continued through the weekend.

By April 30, new targets, including media Web sites, came under attack from electronic cudgels known as botnets. Roughly 1 million unwitting computers worldwide were employed. Officials said they traced bots to countries as dissimilar as the United States, China, Vietnam, Egypt and Peru.

By May 1, Estonian Internet service providers had come under sustained attack.

On May 9 a new wave of attacks began at midnight Moscow time. By his account, 4 million packets of data per second, every second for 24 hours, bombarded a host of targets that day.

By May 10, bots were probing for weaknesses in Estonian banks and especially in credit-card verification systems. They forced Estonia's largest bank to shut down online services for all customers for an hour and a half.

In the end, Estonia was unable to effectively counter the attack. It cut its Internet connections to the outside world so that people within Estonia could continue to use their conventional services.

B. *Patterns of a Cyber War*

Although other Cyber Wars, like for instance the so called Cyber War 2.0 (Russia vs. Georgia), differ in detail (first and foremost by the correlation between Cyber War and traditional "Kinetic War"; "standalone Cyber War" in Web War 1 vs. Cyber War parallel to Kinetic War at Cyber War 2.0 1), the following general patterns of a Cyber War can be identified:

1) Cyber Espionage

First in every war - as Sun Tzu puts it - 'you need to know your enemy and yourself'. Long before the actual Cyber Attack, usually done through a long period

of time, you need to obtain as many secrets as possible from your opponent (for example done with social network analysis, sniffering, conventional spies etc.).
With regard to Cyber War in Estonia, the attackers need to know potential targets (like for instance the web address of the Parliament and government, important Estonian banks and their credit-card verification systems etc.).

2) Preparation of the battlefield

Still within peacetime and after you know your opponents strength and weaknesses, it's time to prepare the battlefield. Once you made the decision, to go on war (even if you need to do it only potentionally), it's time to bring the troops in a good initial setting. In the case of Cyber War it's also time to choose new weapons. This usually comprises not only port-scans, placement of logic bombs or trapdoors etc. but also new kinds of weapons. Stuxnet is a good example for these new and highly specialized weapons targeting on specific industrial equipment [9].
Concerning the example above, attackers need to (i) know the versions of the web-servers and potential exploits in order to do the defacing of the web-site (ii) have the ability to use a Botnet during the war-time or (iii) know weaknesses in Estonian banks and credit-card verification systems etc.

3) Cyber Attack

Now it's time for the "*hot Cyber War*". Like in traditional, kinetic wars, strategy and especially the timing can be crucial. A good timing will obviously have a positive effect on the attack results. The defender will always try to limited attack opportunities and thus defend attacked services with the aim to finally win the Cyber War. This goal is hard to achieve since quick results are difficult to accomplish. The defender will try to win time thus the attacker has to consume more and more resources to continue his Cyber War attack.
In terms of our example, the first targets were governmental Web sites, followed by online news portals and ISPs and finally financial services. In the end Estonia had to cut down their Internet access to the rest of the world. This extreme action which was needed to defend the Cyber Attack shows that such an attack may have an extreme impact on the Internet connectivity of a whole region.
Several defence actions are possible, in the next section will discuss these defence opportunities.

IV. CYBER DEFENCE

When it goes to defence, one of the basic cornerstones, before making a decision, especially when it goes to distributed collaboration, is the *situational awareness*.
The term situational awareness is used frequently in computer science, further we will use the following definition:
"The perception of the elements in the environment within a volume of time and space, the comprehensions of their meaning, and the projection of their status in the near future" [10], [11].
In analogy to chess: Before you can perform a reasonable turn, you need to know the position of (preferably) all chessmen. Without a clear understanding of the current situation, you are not able to choose the winning strategy.

With regard to Cyber War, situational awareness gives answers to questions like:

- Is a Cyber War taking place right now/about do begin (when?)
- Who is attacking?
- What is being attacked?
- What kinds of methods are used for the attack?
- etc.

The sooner an attack (or the intention of an attack) is detected, the better are the defence instruments (more time to identify the real attacker, less systems affected, less blurred traces ...). But to achieve situational awareness in the area of Cyber War is not as easy as it sounds. Up to now, detecting whether a nation is engaging in *Cyber Espionage* (step 1) or *Preparing the battlefield* (step 2) is close to impossible (mainly because of the long period of time in which the actions are taken) [4]. Even within the third step, *the Cyber Attack*, simple things like attack paths or the actual attackers are hardly recognized or identified. Situational awareness in the field of Cyber War is still not sufficiently solved and an open research problem.
Referring to this, John Arquilla has argued that:
Cyber War is like Carl Sandburg's fog. It comes in on little cat feet, and it's hardly noticed. [12].
It's time to close this gap. We need to be able to inform of a future danger in order to prepare for the danger and act accordingly to mitigate or avoid it. That's why so-called EWS are - especially in the field of Cyber Defence - of very high importance.

V. IMPORTANCE OF EARLY WARNING SYSTEMS ON CYBER DEFENCE

The increasing importance of EWS is manifested in the enormous number of various research initiatives around the world such as GENI and FIND in USA, FIRE, OneLab, AutoI in Europe, NWGN in Japan, and FIF in Korea. Trillions of devices, petabytes of data, gigabytes of transfer speed, payload of packets encrypted, IPv6 as well as the virtualization of services and data impose high requirements on developing a proactive action of the Internet of the Future. Key challenges in such a highly complex environment where data and services are also located somewhere in "clouds" are *security, privacy and trust* [13].
Traditional network-based intrusion detection or intrusion prevention approaches cannot cope with such challenges. The need to protect the infrastructure of the Internet of the Future, as well as to manage such a security infrastructure has to have the highest priority. As stated by ENISA (European Network and Information Security Agency [14]), privacy and trust in a network world are already nowadays the basis for using the Internet for business, communications, social networking etc. In the Internet of the Future, where (i) all devices communicate among each other, (ii) a seamless integration of networks enables the end user to "see" only

"one network", (iii) the data and services are located or are provided somewhere in the "cloud", security, trust and privacy is needed. As traditional approaches are not sufficient any more, we need something completely new to proactively protect the infrastructure of the Internet of the Future and manage these security mechanisms in a consistent manner. More precisely, it is necessary to address the following research issues:
If we assume that the payload of packets will be encrypted because of privacy and security requirements, and also because of the huge amount of data flows, it is not possible to perform deep packet inspection. Therefore, the question that arises is what data features are exploitable for detecting an attack or a deviation from the normal network behavior?
The resulting objectives are therefore:

- An analysis and evaluation of available data (e.g. flow information, sensor data), according to developed evaluation criteria and with respect to the relevance to detect a potential attack resp. a deviation of normal behavior. Hereby an analysis of passive and active measurement techniques and possibilities is necessary in addition to the relevance of the available data.
- Development and application of correlation techniques (e.g., temporal correlation, topological correlation) of various data sources, development and application of AI approaches.
- Development of methods for trend analysis in risk management.
- Modeling of network behavior, identification of the deviation from "normal" behavior.
- Determine EWS data sharing.
- Cooperative behavior, EWS have to be able to form binding commitments.

If we assume that data and services will be located, resp. provided in "clouds", then the architecture of an EWS must address this virtualization aspect. Thus, virtualized security architecture is needed. Although virtualization is a mainstream technology nowadays, it seems that security issues are often an afterthought. Existing security models and practices cannot be directly applied to a vastly different environment. Furthermore, virtualization principles could change drastically the way we do security, that forces to rethink how to manage these security items. If we assume a pervasive environment, it is necessary to develop and adapt machine learning techniques to cope with new challenges and the changing environment.
The *objectives* of an EWS are (i) to protect next-generation networks by developing a *sophisticated next-generation EWS*, (ii) to develop novel architectures, sophisticated models for network behavioral analysis and learning algorithms in order to build the next-generation EWS, able to deal with specifics such as encrypted payload of packets, trillions of devices and petabytes of data as well as IPv6 networks, and (iii) to develop approaches and models to define "normal" behavior and anomalies, threat levels, EWS data sharing. Finally raising

management aspects have to be solved taking latest overall security management initiatives in mind [15].
Key elements can be summarized as follows:

- Protect next-generation networks by developing a *sophisticated next-generation EWS*
- Develop novel architectures, sophisticated models for network behavioral analysis and learning algorithms in order to build the next-generation European EWS system, able to deal with specifics such as encrypted payload of packets, trillions of devices and petabytes of data as well as IPv6 networks
- Develop sophisticated correlation approaches to analyze, correlate and evaluate existing data (e.g. flow information), and to reason about threat levels on basis of existing data; develop novel methods of detecting malware-driven network beaconing and command & control channels using both temporal and spatial flow attributes; develop concepts of fuzzy searching for resilient and adaptable malware detection at various sensing points in the network; investigate techniques for interpreting distributed sensor data for broader situational awareness
- Fundamentally improve the state-of-the-art in automated network threat blacklist derivation
- Develop approaches and models to define "normal" behavior and anomalies, threat levels, EWS data sharing; improve the automated assimilation of new security advisories and early warnings
- Improve the understanding of virtualization security (e.g., virtual sensors), develop new security models

Based on the requirements, we have evaluated current approaches and in the next section, we give a comparison of the different technologies.

VI. OVERVIEW OF EXISTING SYSTEMS AND APPROACHES

Currently, two fundamental techniques are used for network-based intrusion detection: misuse detection and anomaly detection.
The first one encompasses the signature based group of systems. There, the detection is accomplished by the definition of malicious behavior i.e. by a set of patterns saved in advanced and discarded in a database for example. The traffic is checked for the presence of a previously known pattern either by testing the whole packet including the payload or by simply checking the header. This is the most common type of intrusion detection system (IDS) and widely in use. Especially in an environment with very high bandwidth it is not possible to inspect the complete payload because of the amount of processing power needed. Some approaches try to overcome these restrictions by applying machine learning techniques to achieve a complete payload inspection with bandwidth over 1 Gbps, for example the project ReMIND from Fraunhofer [16].

Systems that are based on anomaly detection construct a behavioral model that describes the positive behavior, thus the types, amounts, daily traffic allocation, etc. of the monitored network. The detection is realized by the measurement of the current state of the system and the comparison to the values gained from the model. Therefore, machine learning techniques are used, such as expert systems, data mining algorithms, evolutionary algorithms, and neural networks.
Also combinations are possible, for example the application of Evolutionary Algorithms in Data Mining Systems. This type of IDS is also called Network Behavioral Analysis (NBA). An enhancement of NBA systems is called Network Situation Awareness (NSA), where visualization and high-level data management are included to the process of network monitoring.
The main challenges for the protection of the network and the detection of malicious traffic and behavior comprise among other things are data and alarm correlation, source determination and forensic capabilities.
Current techniques to address these requirements are routers as honeypots, DDoS detection with honeypots, traffic diversion to honeyfarms, other information sources (like system, security and network capture/trace data), usage of protocols (like BGP, MPLS, Netflow, etc.), and usage of the human eye to catch anomalies. Due to increasing bandwidth and increasing number of services, the current systems are already hardly able to keep up with the development, and further systems will not be manageable anymore.
In the next section we will shortly describe the work of related projects. These projects try to find solutions for current Internet-related problems; early warning systems especially focusing on security aspects of the Internet of the Future.

A. *Early Warning and Intrusion Detection based on Combined AI Methods*

The project Early Warning and Intrusion Detection System Based on Combined AI Methods (FIDeS) funded by the German Ministry of Research and Education (BMBF) aims at developing an advanced, intelligent assistance system for detecting attacks from the Internet both in local area networks and in wide area networks as early as possible [18]. Within the framework, widely-used Internet protocols such as FTP, SMTP, and HTTP shall be considered, but also newer protocols such as SOAP. In addition, fraudulent access in security-critical, IT-based business processes of enterprises will be detected. Conventional IDS and in particular IDS for anomaly detection usually produce a high false positive rate or do not detect all attacks (false negatives). Complementary to anomaly-based IDS, the project develops an early warning system based upon using different methods of Artificial Intelligence (AI). This system supports a security officer in analyzing attacks and carrying out appropriate counter measures. Consequently, the project FIDeS focuses more on assistance (such as concrete instructions in case of an attack) rather than on mere intrusion detection. For this purpose, various AI-based methods are used such as declarative knowledge representation, the generation of explanations, and cognitive assistance.

B. *Event Monitoring Enabling Responses to Anomalous Live Disturbances (EMERALD)*

The EMERALD (Event Monitoring Enabling Responses to Anomalous Live Disturbances) environment is a distributed scalable tool suite for tracking malicious activity through and across large networks [19]. EMERALD introduces a highly distributed, building-block approach to network surveillance, attack isolation, and automated response. It combines models from research in distributed high-volume event correlation methodologies with over a decade of intrusion detection research and engineering experience. The approach is novel in its use of highly distributed, independently tunable, surveillance and response monitors that are deployable polymorphically at various abstract layers in a large network.

C. *ARAKIS*

ARAKIS [20] is a nationwide, near real-time, network security event early warning system developed by NASK and operated by CERT Polskai [21]. The system consists of a central repository and distributed sensors that collect and correlate data from different sources including low-interaction honeypots, firewalls, anti-virus systems and darknets. The system is oriented towards detection and characterization of new attacks based on the automated analysis of captured honeypot payloads and supporting data. Further information can be found at [21].

D. *WOMBAT – Worldwide Observatory of Malicious Behaviors and Attack Threats*

WOMBAT is an European project (STREP) under the FP7 ICTWork Program 2007–08 Objective 1.4 [22]. The project aims at providing new means to understand the existing and emerging threats that are targeting the Internet economy and the net citizens. To reach this goal, the project is structured around the three following main objectives:

1) *Real time gathering of a diverse set of security related raw data*
2) *Data enrichment by means of various analysis techniques*
3) *Threats Analysis*

The acquired knowledge will be shared with all interested security actors (ISPs, CERTs, security vendors, etc.), enabling them to make sound security investment decisions and to focus on the most dangerous activities first. The project also aims to increase the level of confidence of European citizens into the net economy. Project results and innovation are new data gathering tools, advanced features (high interaction, real-time analysis), new targets (802.11p (car-to-x), bluetooth, RFID, etc.), tools and techniques for characterization of malware, and malware-based analysis and contextual analysis.
The WOMBAT project has shown so far that the generation of good benchmarks for malware detection techniques is a challenging problem, especially:

- Amount and dynamics of nowadays malware makes the generation of an exhaustive sample set an almost impossible task
- Importance of filtering samples to spot cases that could potentially lead to ambiguities
- Problem of labeling: how to define whether the label assigned to a sample is correct?

VII. WEAKNESSES OF THE CURRENT APPROACHES

Although there are some components which try to monitor the status of the Internet and to detect new threats and network anomalies; these systems suffer from the following shortcomings:

- Internet telescopes and monitoring systems **strongly rely on the use of the dark address space**. Although this is efficient for the detection of worms, network scans, etc., target-oriented attacks are hard to be recognized [23].
- Misuse detection is realized in particular by means of **Deep Packet Inspection (DPI)** and the evaluation of header information. DPI does not scale well with massive bandwidth levels, such as those at the Internet backbone [24].
- One of the most important sources for information is the evaluation of **flow data**. All of the systems in use strongly rely on the evaluation of sFlow which is a sampling technology and therefore not able to provide 100% accurate results [25].
- Early Warning Systems only evaluate **logs, flows** or are realized by **packet counting** [26].
- The inherent division between network and host-based indicators is a weakness of the current approaches. Currently, there is no robust system known that **effectively correlate** these disparate data streams [27].
- **Anomaly detection** is only realized **in subnets** and it is extremely difficult to profile "normal" behavior with any level of identity [24].
- The operation on an **inhomogeneous and non-interoperable** security infrastructure, containing stovepipe systems, and application- and task-specific "security silos" is a shortcoming of state of the art approaches [28].

In the context of the Internet of the Future, the difficulties to adapt these systems are even worse because of the changing characteristics (illustrated in Figure 1). An overview of the requirements for current approaches is given in Table I.

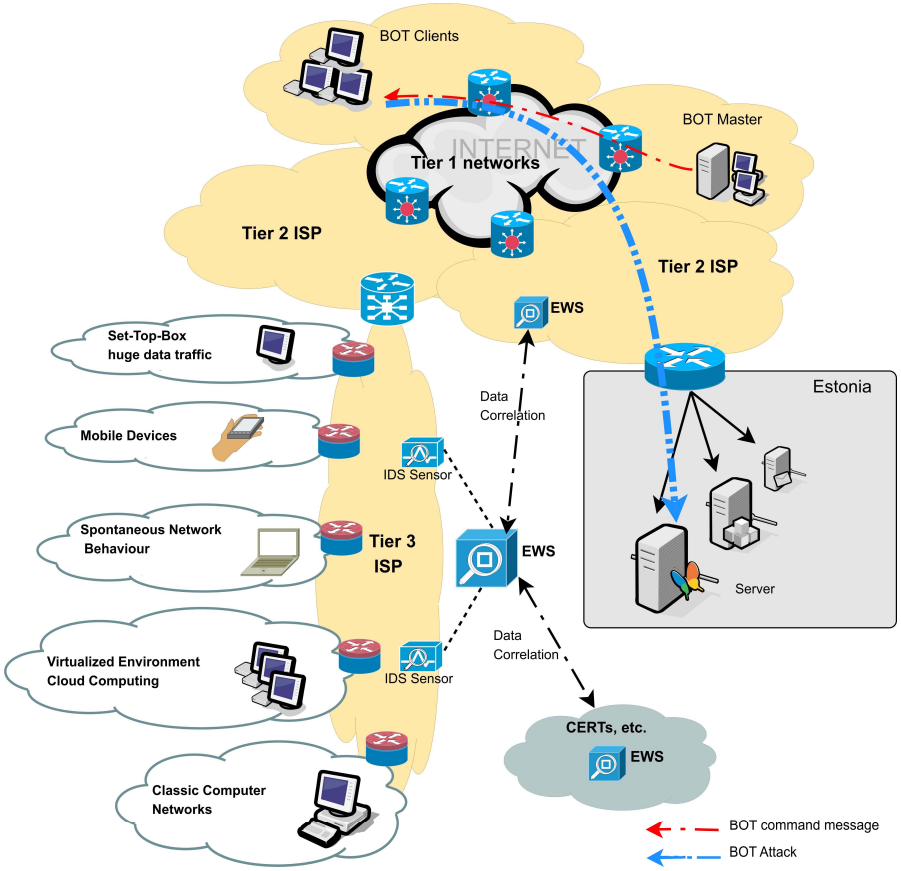

FIGURE I CHALLENGES FOR FUTURE EWS TECHNOLOGIES

TABLE I CAPABILITIES OF STATE-OF-THE-ART EWS TECHNOLOGIES

Requirements	Misuse Detection	Anomaly Detection	NEWS plugin [36]	A-EWS [17]	SANS ISC [37]	ATLAS threat index [38]	FIDeS [18]	EMERALD [19]	ARAKIS [20]	WOMBAT [22]
Extended Flow Handling	X									
Sophisticated Correlation of data				(X)			(X)	(X)		(X)
Comprehensive Reasoning Model										
Traffic Volume Independency	X	X		X	X	X	X	X	X	(X)
End-System Independency	X	X		X	X	X	X	X	X	(X)
Payload-independent analysis		(X)	(X)	(X)						
Safeguarding Mobile Devices										
Virtualized Environment / Clouds	X									
Spontaneous Network behavior										

Thus, the following requirements have to be fulfilled:

- **Extended Flow Handling:** As proposed by CISCO Visual Networking Index the global mobile data traffic will increase 26-fold between 2010 and 2015 reaching 6.3 Exabyte per month [29], [30]. Also, CISCO assumes that global IP traffic will exceed to about 767 Exabyte from 2009 to 2014. Traditional signature based approaches are not sufficient to that global IP traffic. Obviously, efficient signature flow solutions are needed in future EWS. Thus, extended flow handling is required to detect malicious traffic in the future [31].

- **Sophisticated Correlation of data**: Not only data collection algorithms are required, without data correlation an EWS will be unable to detect events. In future networks different data sources (internal IDS, ISP-CERT, national CERT, etc.) have to be analyzed with different analysis methods (signature, anomaly, etc.). Thus, a sophisticated correlation of data of different sources and methods is a requirement for an EWS [32].

- **Comprehensive Reasoning Model**: Current approaches used in intrusion detection systems are based on the traditional views of computer security. An alternative view that may provide better security systems is based on adopting the design principles from the natural immune systems, which in essence solve similar types of problems in living organisms [33]. Artificial immunology concepts for handling intrusion detection through approximate reasoning have to be used in future EWS.

- **Traffic Volume Independency & End-System Independency**: Scalability, such as independency of data traffic is needed to sufficiently detect intrusions in huge computer networks. Detecting malicious network behavior should be independent from IP traffic generated by different end-systems.

- **Payload-independent analysis:** In future networks the amount of connected devices will increase dramatically [34] next to global data traffic [29], [30]. Payload analysis may not influence the EWS detection of malicious traffic.

- **Safeguarding Mobile Devices**: The amount of global data traffic from mobile device will continue to increase significantly [30]. Future EWS have to be aware of mobile data traffic.

- **Virtualized Environment / Clouds**: Currently the usage of cloud services are widely discussed, in the near future cloud services will be largely used. On the one hand future EWS could benefit from this concept by usage of cloud services on the other hand EWS have to cope with new kind of attacks on these new services.

- **Spontaneous Network behavior**: As proposed by CISCO the number of mobile device will increase and therefore the proportion of ad hoc based traffic will increase in the following years [30], [35]. Therefore, future EWS have to cope with spontaneous network behavior.

VIII. EARLY WARNING IN THE FUTURE

The inter-relationships and inter-dependencies between formerly stand-alone systems and networks are leading to complexities in the infrastructures of our society that have never been seen before. These complex systems and networks disseminate and process massive amounts of personal and business data, information and content in ways which are difficult to understand and control for users, in particular private citizens. In recent years we have witnessed a growing series of accidents and attacks on the Internet and on applications and databases. Through denial of service attacks, viruses, phishing, spyware and other malware, criminals disrupt service provisioning and steal personal or confidential business data for financial gain or other purposes. An increasingly organized and efficient though disruptive e-market is thus taking shape on an international scale. Although we do not know how the Internet of the Future will look like, some characteristics can be identified:

- Layered, but augmented by a number of cross-cutting dependencies.
- Multitude in scale compared to the current Internet, billions of entities including things.
- Spontaneous and emerging behaviors and unanticipated new usages.
- Trust, privacy and security as key components.
- Pervasive digital environment, heterogeneous infrastructures, terminals and technologies.
- User-centricity and usability is critical.
- Enabling the "Internet of Services" and its new business models.
- Trust, privacy and security as key components, managed according a common security policy.

In respect to these characteristics the aim of our requirements is the development of an efficient cooperative Early Warning System for future networks.
In the current environment of the Internet, multiple distributed and heterogeneous networks are connected at which no encryption is done or mostly only partial. Security-related cooperation between autonomous system providers is only done on a very marginal level. Anomaly detection, which is a very powerful instrument for intrusion detection, is only possible and available for subnetworks, while current EWS are based on the analysis of log-files, flow-information or packet counting. Characteristics of the Future Internet will include a virtualized environment, IPv6 network, continuous payload encryption, an enormous number of devices and data as well as a highly distributed and pervasive environment. Therefore, most of the current components and management approaches are not applicable or sufficient any longer. To overcome these shortcomings, an efficient EWS has to be based on a network virtualization and will implement an EWS based on the use of virtual sensors, new reasoning models, new developed learning

algorithms and a sophisticated correlation of data also taking into account security management aspects.
A long-overdue EWS will help the region to avoid deliberate or inadvertent outages, reduce the spread of new computer malware, and ensure continuity of services. Furthermore, the Future Internet has no centralized control hub and its complexity is not bounded by geographical, political, administrative or cultural borders. EWSs are present in various systems and are a crucial component of effective risk management in enterprises and for national homeland security systems. An Internet-wide EWS however is still missing.
Because of the identified characteristics of the Internet of the Future, an EWS has to overcome the following issues that make the use of current State-of-the-Art Intrusion Detection Systems impossible or disadvantageous:

- **Applicability**: The **persistent payload encryption** blights Deep Packet Inspection.
- **Computational effort**: High bandwidth, numerous, **highly dynamic connections and huge amounts of data** would necessitate enormous amounts of computational power for deploying traditional systems and algorithms.
- **Energy consumption**: Mobile devices are becoming more and more important and the mobility will be one of the main characteristics of the Future Internet. Because of the increasing complexity of mobile applications, the processing capabilities of the mobile hardware and the endurance of the battery, it is neither possible nor desirable to set up sophisticated IDS on these devices. Therefore, the protection of the whole network from inside out is necessary and thus the **intelligence has to be brought back to the network**.
- **Novel threats: threats and attack possibilities** evolving from the highly dynamic environment with billions of devices cannot be handled by current systems, are not even known today.

IX. CONCLUSIONS

The Internet of the Future will consist of dynamically scalable and virtualized resources, which will be provided by providers as a service over the Internet. 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 Internet of the Future will need to cope with new problems.
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 handle the remaining old ones. They have to cope with attacks on their network, but their well-established *Intrusion Detection Systems* (IDS) and *Early Warning Systems* (EWS) will not defend them anymore, because the packet payload will be encrypted. As all current IDS and EWS installed by the providers rely on analyzing the packet payload or packet

headers to proper fulfill their tasks, we need a new kind of EWS suitable for the needs of future computer networks.

In this paper we have evaluated requirements needed to be fulfilled by an enhanced Early Warning System which is able to inform about ongoing Cyber Attacks. As shown in an analysis, so far no system can completely comply with the requirements presented. An efficient Cyber Defence is only promising if and only if the capabilities and the requirements are congruent as much as possible. Therefore, further research activities are needed in the future to build such enhanced EWS [39].

ACKNOWLEDGMENT

The authors wish to thank the members of the Chair for Communication Systems and Internet Services at the Universität der Bundeswehr München, headed by Prof. Dr. Gabi Dreo Rodosek, for helpful discussions and valuable comments on previous versions of this paper. The Chair is part of the Munich Network Management Team.

REFERENCES

[1] M. Libicki, Cyberdeterrence and Cyberwar. Rand Corporation, 2009.

[2] G. Brown, M. Carlyle, J. Salmer´on, K. Wood, et al., "Defending critical infrastructure," Interfaces, vol. 36, no. 6, pp. 530–544, 2006.

[3] J. Arquilla and D. Ronfeldt, "Cyberwar is coming!," Comparative Strategy, vol. 12, no. 2, pp. 141–165, 1993.

[4] R. Clarke and R. Knake, Cyber War: The Next Threat to National Security and What to Do About It. HarperCollins, 2010.

[5] C. Clausewitz, M. Howard, and P. Paret, On war. Princeton University Press, Princeton, NJ, 1976.

[6] P. Finn, "Cyber assaults on Estonia typify a new battle tactic," Washington Post, vol. 19, 2007.

[7] M. Lesk, "The new front line: Estonia under cyberassault," Security & Privacy, IEEE, vol. 5, no. 4, pp. 76–79, 2007.

[8] "Russian Invasion of Georgia. Russian Cyberwar on Georgia. Report of the Government of Georgia." http://georgiaupdate.gov.ge/doc/10006922/CYBERWAR-

[9] N. Falliere, L. O. Murchu, and E. Chien, "W32.Stuxnet Dossier," tech. rep., Symantic Security Response, October 2010.

[10] M. ENDSLEY, "Toward a Theory of Situation Awareness in Dynamic Systems," HUMAN FACTORS, vol. 37, no. 1, pp. 32–64, 1995.

[11] N. Castellan, Individual and group decision making: current issues. Lawrence Erlbaum, 1993.

[12] "CyberWar! Frontline - Intervies with John Arquilla." http://www.pbs.org/wgbh/pages/frontline/shows/cyberwar/interviews/arquilla.html/.

[13] K. Hwang, S. Kulkareni, and Y. Hu, "Cloud security with virtualized defense and reputation-based trust mangement," Dependable, Autonomic and Secure Computing, 2009. DASC '09. Eighth IEEE International Conference on, pp. 717 – 722, Dec 2009.

[14] "European Network and Information Security Agency." http://www.enisa.europa.eu/.

[15] NATO, "Concept for a NATO Security Management Infrastructure (SMI)," AC/322-D(2008)0049 (INV), Dec 2008.

[16] "ReMIND – Real-Time Machine Learning Intrusion Detection." http://www.first.fraunhofer.de/owx 1 4079 2 2 0 4be4ae1a1c59c1.html.

[17] O. K.-P. Karsten Bsufka and S. Albayrak, “Intelligent Network-Based Early Warning Systems.” http://dx.doi.org/10.1007/11962977, 2006.

[18] “Early Warning and Intrusion Detection based on Combined AI Methods." http://www.fides-security.org/.

[19] “Event Monitoring Enabling Responses to Anomalous Live Disturbances.” http://www.sdl.sri.com/projects/emerald/project.html.

[20] “Arakis dashboard.” http://www.arakis.pl/en/index.html.

[21] “CERT Polska.” www.cert.pl.

[22] “Worldwide Observatory of Malicious Behaviors and Attack Threats.” http://wombat-project.eu/.

[23] A. Shimoda and S. Goto, “Virtual Dark IP for Internet Threat Detection,” in APAN Network Research Workshop, pp. 17–23, 2007.

[24] S. Dharmapurikar, P. Krishnamurthy, T. Sproull, and J. Lockwood, “Deep packet inspection using parallel bloom filters,” in High Performance Interconnects, 2003. Proceedings. 11th Symposium on, pp. 44–51, IEEE, 2003.

[25] “sFlow.” http://www.ams-ix.net/technical/sflow.html.

[26] A. Serjantov and P. Sewell, “Passive-attack analysis for connectionbased anonymity systems,” International Journal of Information Security, vol. 4, no. 3, pp. 172–180, 2005.

[27] K. Wang and S. Stolfo, “Anomalous payload-based network intrusion detection,” in Recent Advances in Intrusion Detection, pp. 203–222, Springer, 2004.

[28] B. Swartout, R. Patil, K. Knight, and T. Russ, “Toward distributed use of large-scale ontologies,” in Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems, 1996.

[29] C. Index, “Global Mobile Data Traffic Forecast Update,” Cisco White Paper[Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf.

[30] C. Index, “Global Mobile Data Traffic Forecast Update, 2009-2014,” White Paper, CISCO Systems Inc, vol. 9, 2010.

[31] P. Laud, “Handling encryption in an analysis for secure information flow,” in Proceedings of the 12th European conference on Programming, pp. 159–173, Springer-Verlag, 2003.

[32] C. Phua, V. Lee, K. Smith, and R. Gayler, “A comprehensive survey of data mining-based fraud detection research,” Arxiv preprint arXiv:1009.6119, 2010.

[33] S. Shahrestani, “Employing artificial immunology and approximate reasoning models for enhanced network intrusion detection,” WSEAS Transactions on Information Science and Applications, vol. 6, no. 2, pp. 190–200, 2009.

[34] F. Mattern and C. Fl¨orkemeier, “Vom internet der computer zum internet der dinge,” Informatik-Spektrum, vol. 33, no. 2, pp. 107–121, 2010.

[35] A. Farooqi and F. Khan, “Intrusion detection systems for wireless sensor networks: A survey,” Communication and Networking, pp. 234–241, 2009.

[36] F. Bustamante and D. Choffnes, “NEWS plugin for the Vuze (formerly Azureus) BitTorrent Client.” http://www.aqualab.cs.northwestern.edu/projects/NEWS.html.

[37] “Internet storm center.” isc.sans.org.

[38] “Active Threat Level Analysis System.” http://atlas.arbor.net/.

[39] T. Guo, “Shaping Preventive Policy in “Cyber War” and Cyber Security: A Pragmatic Approach,” 2011.

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

Towards Next-Generation Intrusion Detection

Robert Koch
Institut für Technische Informatik (ITI)
Universität der Bundeswehr
Munich, Germany
Robert.Koch@UniBw.de

***Abstract*- Today, Intrusion Detection Systems (IDS) are integral components of larger networks. Even so, security incidents are on a day-to-day basis: Numerous data leakage scandals arouse public interest in the recent past and also other attacks like Stuxnet are discussed in the general public. On the one side, the commercial success of the Internet and the possibilities to carry out attacks from a relatively safe distance attracts criminals and made e-Crime to a multi-billion dollar market over the past years. On the other side, more and more services and systems migrate to the Internet, for example Voice over IP (VoIP) or Video on Demand (VoD). This enables new and potential attack vectors.**

With the steadily increasing use of encryption technology, State-of-the-Art Intrusion- as well as Extrusion Detection technologies can hardly safeguard current networks to the full extend. Furthermore, they are not able to cope with the arising challenges of the fast growing network environments.

The paper gives an overview of up-to-date security systems and investigates their shortcomings. Latest security-related threats and upcoming challenges are analyzed.

In the end, requirements for a Next-Generation IDS are identified and current research as well as open issues are presented.

Keywords: Next-Generation Intrusion Detection, Security Threats, Intrusion Detection, Intrusion Prevention, Data Leakage Prevention, Early Warning

I. INTRODUCTION

With the interconnection of computer systems, numerous security threats emerged. One of the first publications towards IDSs was a technical report in 1980 [1]. A first model of a real-time IDS and a prototype had been built, the Intrusion Detection Expert System IDES [2]. Nowadays, plenty of specialized systems exist, but the basic functionality can be differentiated with regard to the detection technique, misuse- (signature) and anomaly detection (behavior). While the former ones search for well-known patterns, the latter ones build a model of the normal network behavior and attacks can be detected by measuring significant deviation of the current status against the behavior expected from the model. Therefore, anomaly-based systems are able to detect new and yet unknown threats at the cost of higher false alarm rates. The placement of the system, host- or network-based, is another aspect. Host-based systems are able to access a wide range of system information, logs, etc., while network-based systems are only able to evaluate the network traffic. However, because of their installation at central points in the network, they are able to detect attacks against the whole network or distributed attacks, which cannot be detected by a host-based analysis.

Other attributes can be used for a more precise classification, like time-based constraints or the degree of interoperability (e.g., [3, 4]).

Today, well-known attacks or new threats like a worm propagation can be detected and obstructed. Anyway, all systems suffer from important real-world problems. Even more, the current technology trends tighten this situation: Yet available systems will not be able to cope with challenges like encryption or increasing bandwidth.

The further paper is organized as follows: In Section 2, a brief overview of the evolution of security threats is given. Section 3 presents State-of-the-Art security systems and research and points out their most important shortcomings. Based on the identified shortcomings, requirements for Next-Generation Intrusion Detection are derived in Section 4. An architecture of a Next-Generation IDS is proposed in Section 5. Concepts under development, which try to address some of the most important current shortcomings, are presented as well. Finally, Section 6 concludes the paper by highlighting the most important open research issues.

II. THREATS AND TENDENCIES

The scope of attackers and malicious programs has changed significantly over the years. The focus of the first computer virus was on the destruction of data, e.g. formatting the hard disk drive or deleting executable files (e.g. [5]). With the development of worms, automated infection over networks was enabled and used to build botnets, consisting of numerous user PCs without the knowledge of the owners. These networks can consist of hundreds of thousands of infected systems and are used to send Spam or to block services by Distributed Denial-of-Service (DDoS) attacks.

The destructive behavior in the beginning changed to commercial-driven reasons. Today, spammers can participate in Affiliate Programs: They sign up in a program and are provided with an unique identifier. If a sale is backtracked to an identifier, the corresponding spammer is rewarded with a commission [6].
The commercial success of the Internet and the possibilities to carry out attacks from a relatively safe distance attracts criminals and makes e-Crime to a multi-billion dollar market (e.g., see [6, 7]).
Therefore, the profitable and relatively risk-free underground market stimulates the proliferation of malicious code by the creation and selling of attack kits. No technical in-deep knowledge is needed any longer to create new, dangerous malicious software [8]. The first attack kit (Virus Creation Lab, 1992) only provided basic functionality, but state-of-the-art kits like Nukesploit are highly professional and sold for several thousand Dollars. Also different service levels are available, for example unlimited support or regular updates [9]. A major difficulty arising with the professional construction kits are the high numbers of new signatures. A new signature appears with every new created code, building malicious code families.
More and more services migrate to the Internet, for example VoD or VoIP. With more and new services, also more new potential attack possibilities arise. For example, some malicious programs encrypt the data on the infected system and the user has to pay for the key. This type of malicious software (ransomware) appeared for the first time in 1989 [10]. Today, Trojan Horses exists which are able to encrypt data based on public key cryptography [11].
Another aspect is the handling of malicious programs: Latest trends show, that the percentage of targeted attacks continuously increases. E.g., the Hydraq Trojan (Aurora): Several large companies had been compromised by attackers using this Trojan [7]. The attacks started by evaluating data about employees, available on the company's website or in social networks: Social Engineering is on the rise again. Social networks like Facebook or Twitter are in the focus of attackers because of their prosperity of information. Many people are easygoing when dealing with sensitive data in social networks. This information is used by social engineers to create attacks, e.g. Emails with malicious attachments, obviously sent by a friend and with a topic related to the latest movements in the social net. So, the probability that the target opens the attachment and infects the system is very high. Targeted attacks are often constructed for a single or few destinations, so no patterns will be available.
The dissemination routes of malicious software are not restricted to networks: E.g., promotional gifts like USB-sticks given away on trade shows are popular instruments [12]. A Trojan is already installed on the stick. By connecting the stick to a computer, the Trojan installs itself on the system. Therefore, the threat is injected directly onto the target system or network, bypassing the security systems. With the help of offline-propagation, also formerly secure systems and networks like Supervisory Control And Data Acquisition Systems (SCADA) can be compromised. Therefore, a protection against attacks from the outside is not enough.

Data leakage has become an important issue for the last years. In contrast to the insider threat, data leakage includes accidental or unintentional data loss in addition to malicious theft [13]. Numerous scandals about data loss arose public interest, for example see [14, 15]. The insider threat is one of the most challenging endangerments today. While governments and the military had been in the spotlight of attacks during the cold war, today the industry is the most important target for espionage. A recent study specified the economic loss for each individual business company in Germany on an average of about 5,57 million Euro in 2009. 61 % of all large-scale enterprises had been hit by business crime in the past two years [16].

The particular endangerment by the insider is based on the authorized access and the knowledge about the security mechanisms. Also, by the widely spread use of data storage mediums like memory sticks, it can be easy for a legitimate employee to extrude confidential data if no protection mechanisms are in place. The released numbers of the percentage of the insider threat compared to all incidents of data loss differ keenly and go up to 80 % and more. The Verizon Data Breach Investigation Report attracted interest in 2008, because their evaluation of the insider threat presented a value of only 18 % [17]. Anyway, in the Report of 2010, Verizon published a proportion of about 48 % incidents caused by insiders after evaluating a wider range of cases [18]. In addition, the estimated numbers of unreported cases based on insider jobs are much higher, because numerous companies do not press charges because of a possible loss of reputation. The detection of data leakage is difficult by nature, but the situation is even worse, because high damage only can be avoided by immediate reactions. Beside this challenges, the technical evolution of the Internet opens up additional problems. More and more services are offering protected access. For example, the well-known Firesheep [19] addon for the Firefox-browser attracted numerous people. It enables easily operated HTTP session hijacking attacks. While these security hole existed for several years without concerning the public interest (because of the complex way to utilize it), the addon is easy to use [20]. Therefore, anybody is able to take over a foreign session. The tool comes with filters for e.g. Facebook, Twitter and GMail. So, numerous services like Facebook announced to switch their services to TLS. The trend towards the use of encryption will also be enforced with the broader application of IPv6 as IPSec is a mandatory component of IPv6 [21]. In February 2011, the last address blocks of IPv4 had been assigned. This should speed-up the utilization of IPv6 in the near future; at the moment, less than 1 percent of all traffic is IPv6 (e.g., [22]). Encryption can train the application of IDSs, therefore being a crucial factor.

Important is the shifting from attacks directed onto the operating systems or network protocols to attacks of vulnerabilities in the application layer. The nonstop evolution of the applications results in complex programs and flawed program code. Today, over 70 % of all attacks are targeting the application layer [7]. Most utilized vulnerabilities are provided by browsers and programs like the Acrobat Reader (e.g., [23]). Based on that, the number of Zero Day vulnerabilities increased in recent years.

Vendors are sometimes delaying patches unnecessarily by using a fixed patch-day policy: Program updates are only published on a regular basis (e.g., [24]). Also the safety awareness of the users is inadequate, many users are overstrained by complex and often changing security mechanisms and program configurations: The most successful exploits are taking advantage of vulnerabilities first reported more than a year ago [7].
Current available system are hardly able to cope with these trends. Summarized, the following threats and tendencies are identifiable and emerging:

1. New and yet unknown attacks (new services, devices, etc.)
2. Increasing number of Zero Days
3. Social Engineering and targeted attacks
4. Exploitation of vulnerabilities in the application layer
5. Increasing insider threat
6. Risk of data leakage
7. Ascending use of encryption technology
8. Users are negligent with security-related tasks

Following, current IDSs and techniques will be considered with respect to these properties.

III. CURRENT SYSTEMS AND SHORTCOMINGS

SNORT [25] is a signature-based Network-IDS (NIDS) and Intrusion Prevention System (IPS) capable of performing real-time traffic analysis and packet logging. To gain acceptable results regarding the false alarm rates, signature-based systems like SNORT have to be configured strongly depending on the hosts and services presented in the network. If the system generates many false alarms, no administrator will pay attention to the IDS after a few days. However, a complete in-depth configuration of all systems and services is time-consuming and difficult. Also, the configuration has to be administered all the time: Small changes like an update can have a significant impact. Therefore, the application of signature-based techniques in big network environments often is not successful.
Signature-based systems are reactive by nature [26] and restricted to already known attacks. Therefore, the efficiency of an IDS relies on the update-rates and response-times of the responsible company.
For example, on January, 20th 2011, the latest IDS signatures of Juniper and Sourcefire were released on January, 18th while the latest signatures of Proventia and IntruShield were published on January, 11th (as seen on [27]). Therefore, the up-to-dateness of the signature databases is a crucial point. Lippmann analyzed the effect of identifying vulnerabilities and patching software with regard to IDSs and the up-to-dateness of their signatures [28]. The signatures for IDSs are often not faster available than the publication of software-patches. However, vendors like Microsoft or Adobe often use patch-days for publishing numerous patches at once, therefore delaying patches unnecessarily. For example, latest threats opened up by the vulnerabilities in the Internet Explorer [29] have not been fixed in the

consecutive patch-day even exploits had been published meanwhile and an easy deployable code had been included in the Metasploit-Framework [30].
Anomaly-based systems do not need a signature database, instead they use a model for the evaluation. The accurate modeling of network behavior is an active field of research [31]. The difficulty of behavior-based models is the possibility of misinterpretation of permitted but unknown legal user actions, resulting in high false alert rates. Often, a learning phase is needed to train the corresponding detection model. Online and offline learning must be differentiated. The former one is an incremental respectively sequential training: Learning is performed piece-by-piece in a serial fashion on one individually (randomly) selected training sample set. The latter one takes the whole problem data in one learning iteration [32]. All data has to be labeled in advance due to its benign or malicious character which can be a difficult task (e.g., see [33, 34]). Because of the time and effort needed, Almgren and Jonsson use active learning methods to reduce the needed amount of labeled training data [35].
Even more, numerous anomaly-based systems use online learning in the productive environment (e.g., [36], [37]). Attacks can take place or malicious code can already be in the network during the learning phase, resulting in an erroneous model [33]. Therefore, the system will recognize the previously learned attacks as normal behavior and no alarm will be raised [38, 39]. A possible countermeasure is the use of malicious rather than benign data for the training. Winter et al. proposed a one-class support vector machine (SVM) for the analysis. The system is trained with malicious network traffic [40]. Anyway, also the usage of negative behavior is difficult in matters of completeness. To countervail the endangerment of a learning phase, using unsupervised learning methods can be a solution. Numerous machine-learning approaches have been developed over the last decades. Examples are statistic-based systems, data mining, expert systems, supervised learning-based approaches like neural networks and unsupervised learning-based approaches like k-means- or self-organizing feature maps (SOM) ([38, 41, 42]). Sometimes, supervised and unsupervised concepts are combined (e.g., see [43]). Casas et al. proposed a robust clustering technique to detect anomalous traffic flows based on a sequentially captured temporal sliding-window basis [44]. The approach is completely unsupervised and able to detect attacks without relying on signatures, labeled traffic or training. The system can be directly plugged-in and starts working from scratch without previous knowledge.
Signature-based systems are using string matching techniques to find known patterns of malicious code. This is a computational complex task and can generate up to 80 percent of the total processing time of the IDS [45]. Payload filtering delay has become the main cause of the reduction of network performance [46]. Because of that, software-based NIDSs are hardly able to keep up with traffic over 200 Mbps [47] when executing a full payload analysis (Deep Packet Inspection (DPI)). Therefore, numerous designs and algorithms for hardware-based acceleration have been proposed in recent years. Today, two categories of hardware approaches can be classified, logic and memory architectures. Logic architectures mostly use on-chip logic resources of Field-Programmable Gate Arrays (FPGA) to convert regular expression patterns into parallel state machines

or combinatorial circuits. Memory architectures compile string patterns to finite-state machines and store the corresponding state transition tables in memory [48]. Therefore, memory architectures are more flexible because they allow on-the-fly pattern update without resynthesis and relayout which is needed by logic architectures. By the use of FPGA, the string matching process can be accelerated strongly (e.g., see [47, 49, 50]). On the other side, a tremendous amount of chip resources is used by the growing rule sizes. A lot of work is done to reduce the required number of logic cells per search character (e.g., see [45, 51]). Kong et al. argue that all on-chip solutions are not scalable in long term [46]. Memory and on-board architectures are more likely able to keep up with the increasing set of rules and bandwidth. For example, pre-filtering [47] or prefix and suffix sharing for the rules [46] enables higher speeds (6.4 Gbps respectively 4 Gbps). Even FPGAs can reach high throughput speeds, Gao argues that already light-weight systems like Snort cost too many hardware resources [52]. Also, with the growth of the signature sets and the design scale, the interconnecting latency increases. Therefore, the operation clock frequency and throughput speed will drop. Gao also mentions, that a larger design requires more time for updating and reconfiguration procedures. During that time, the system is vulnerable, e.g. for a new worm spreading. Therefore, Gao uses Ternary Content-Addressable Memories (TCAM) for signature matching, reaching 2 Gbps (and theoretical much higher values) for the Snort signature set.

Other approaches use the network interface card (NIC) for implementing efficient and fast detection systems. While Otey et al. only use header information for the evaluation [53], Bruijn et al. analyze different levels of abstraction, e.g. packets, streams and aggregates in their system [54].

If the mounds of data are to high for a payload or at least a header inspection, Flow-based evaluation can be fulfilled. IPFIX (RFC 5472 [55]) defines a Flow as a group of packets that share a common set of properties. The Flow is completely specified by that set of values, together with a termination criterion (like inactivity timeout). Two important standards are NetFlow ([56] et al.) and sFlow [57]. Plenty of tools are available for the evaluation, e.g. Scrutinizer [58] or NetFlow Analyzer Professional [59]. Flow-based evaluation enables the possibility to analyze higher bandwidth. However, using Flows introduces a delay, therefore the system is not able to initiate fast, near real-time countermeasures [60]. Even more, attacks on the application layer (which are already the most important attacks and still rise) often cannot be detected by the evaluation of the Flow parameters. Figure 1 gives an overview of the implications and correlations of important threats and technology trends for anomaly- and signature-based systems.

In addition to IDSs, the area of Data Leakage Protection (DLP) and Extrusion Detection Systems (EDS) is an emerging sector in recent years. While IDSs focus on the attacker and malware trying to enter and infect the systems from outside, EDSs are monitoring the outgoing traffic searching for keywords or verifying the compliance of the communication with the policies of the company.

When coping with data leakage, DLP systems can cover up to three areas depending on the functionality, namely data at rest (databases, files, etc.), data in

motion (network traffic) and data in use (data traveling to peripherals like DVD burner or printer) [61]. Therefore, packet inspection, session monitoring, encryption and other techniques are used by a DLP.

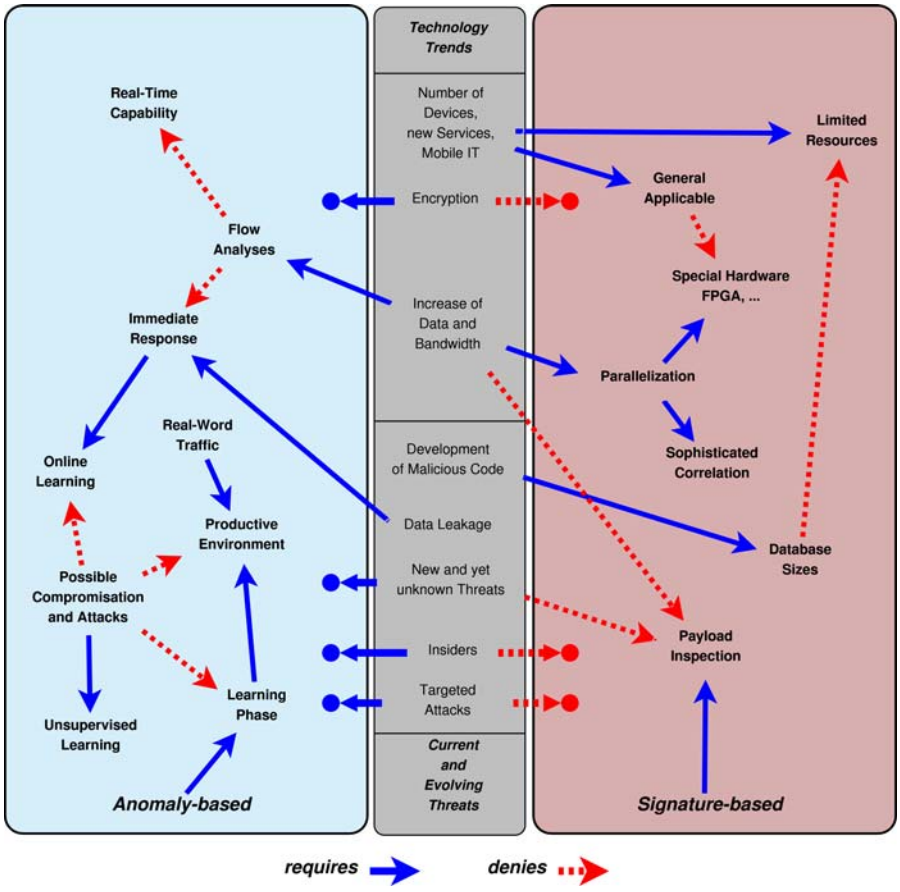

Figure 1. Evolving threats and technology trends and consequences for signature- respectively anomaly-based IDSs.

In the area of DLP, the first implementation was Security Enhanced Linux (SELinux) developed by the National Information Assurance Research Laboratory of the US National Security Agency (NSA), Red Hat and some other companies [62]. SELinux was released as Open Source in 2000 and is included in several Linux distributions today. Plenty other DLP systems and services are available recently, for example from Trend Micro [63] or IBM [64].

Additional to the IDSs, Early Warning Systems (EWS) are in operation. Compared to IDSs, EWSs are larger scaled, monitoring data sources distributed over the whole Internet. The information of attacks in one subnet can be used to alert and

safeguard other sub-networks, which are not yet under attack. For example, the spread of a new worm can be easily detected in the early phase of its run.
In 2001, the Internet Storm Center (ISC) of the SANS Institute was established [65]. It is an analysis and warning service for the Internet and consists of sensors covering 500.000 IP addresses around the globe. Firewall and intrusion detection log entries are collected and sent to the DShield database of the ISC. Abnormal trends and behavior is identified through human volunteers and automated evaluation. Based on that, the handler on duty sets the Infocon level which should reflect changes in malicious traffic and the possibility of disrupted Internet connectivity. Another example is the Arbor Networks Active Threat Level Analysis System (ATLAS) [66].
The NEWS (Network Early Warning System) plugin [67] gathers the collectively provided view of peer computers to detect network anomalies. NEWS uses corroboration from multiple users running in the same area. If enough people see the same problem in the same area, an alarm is raised. The design principles apply to large-scale systems that generate a significant amount of network traffic. Numerous other systems are under development like FIDeS [68] or WOMBAT [69].
The shortcomings and challenges of the current systems are summarized again:

1. Complex configuration
2. Detectability of Zero Days
3. Delays for signature updates
4. Rising bandwidth and data volume
5. Sizes of pattern databases
6. Application-Layer attacks
7. Encrypted network connections

Table I assigns the characteristics to the different systems.

Table I: Shortcomings of current Intrusion Detection / Prevention & Data Leakage Prevention Systems.

	Intrusion Detection						
	Signature-Based		Behavior-Based		Data Leakage		EWS
	Host	Network	Host	Network	Host	Network	*Network*
Configuration	×	×	√	√	×	×	(√)
Zero Days	×	×	√	√	¬	¬	√
Signature Delays	×	×	√	√	¬	¬	√
Bandwidth	×	×	√	(√)	√	(√)	(√)
Database Sizes	×	×	√	√	√	√	(√)
Application Layer	(√)	(√)	(√)	×	√	(√)	√
Encrypted Communication	√	×	√	(√)	√	×	×
Targeted Attacks	×	×	(√)	(√)	¬	¬	×
Distributed Attacks	×	(√)	×	√	¬	¬	√

√ means uncritical while × shows shortcomings of the particular systems, () means restricted applicable, ¬ stands for not applicable. Note that EWS are inherently network-based, therefore there is no column for host-based systems.

IV. REQUIREMENTS FOR A NEXT-GENERATION IDS

Based on the shortcomings of current IDSs defined in Section 3 and the security-related threats and tendencies in the Internet shown in Section 2, the requirements for a Next-Generation IDS are educed. In detail, the IDS has to fulfill the following requirements:

1. Behavior-based analysis: Because of the increasing number of Zero Days, the growth of targeted attacks and the increasing percentage of encrypted communication (benign as well as malicious, e.g. botnet communications), signatures are often not available in time or not possible at all. Even if the application of behavior-based methods is a challenge, sophisticated statistical methods can be used to detect attacks (e.g., see [70], [71]) even in encrypted environments. Other reasons require behavior-based techniques, too: More and more mobile devices like smartphones are participating in the networks. Because of limited computing resources and with regard to the endurance of the batteries, the application of signature-based techniques is not reasonable or even possible. Also in server systems, the necessary near-realtime evaluation of patterns is limited not only by the amount of data to investigate but also by the sizes of the databases and millions of patterns.
2. Abdication of a learning phase: The use of behavior-based techniques often (but not necessarily), requires a learning phase of the system in the productive, real-world environment. Because of the endangerment of the learning phase and the difficult task of creating clean labeled data, this phase must be omitted as far as possible. Unsupervised learning techniques can be used (see Section 3) or the learning phase must be replaced by other techniques. For example, the anomaly-based system developed by Casas et al. [44] does not need signatures, labeled data or training. In the area of neural networks, Moraga examined how to design a neural network only based on knowledge [72]. It is important to understand that the abdication of the learning phase does not transform a behavior-based into a signature-based system: The detection is still fulfilled by the comparison of the measured state of the environment to the prediction of the model.
3. No payload evaluation: For a general applicability the system must be designed without the need of a payload evaluation as far as possible. Even more, the increasing use of encryption denies the use of payload data. Therefore, a Next-Generation IDS cannot rely on the availability of the packet payload.
4. Network-based evaluation and use of agents: Even if a host-based installation has several advantages with regard to the available information (e.g., running processes, decrypted data, log files, etc.), the IDS requires a network-centric design. On the one hand, distributed and sophisticated attacks against the whole network only can be recognized by a network-based installation, on the other side the management of

numerous host-based system is error-prone, complex to manage and often poorly scalable in large environments. Only if it is indispensable, host-based agents should supplement the network-based core system.

5. Cross-evaluation and distribution: The upcoming threats and challenges require an exhaustive use of behavior-based techniques. Therefore, the related false alert rates have to be reduced. By examine ingress traffic and the correlation of anomaly detection alerts of administratively disjoint domains, the false alert rate can be reduced significantly and abnormal data and Zero Days can be detected [73].
6. Active and automated prevention: The system must be able to carry out a completely automated operation. On the one hand, the amounts of data, connections and speed of actions are already too high to be able to permit a reasonable manual interaction. On the other side, especially in the area of DLP, a beginning leakage of data must be stopped as early as possible. The loss of reputation after losing data will often be more expensive (e.g., see [16]) than the costs caused by an misleadingly activated interruption of a single connection. Of course, the probability of a wrongly dropped connection must be very low.

V. ARCHITECTURE OF A NEXT-GENERATION IDS

To fulfill the requirements presented in Section 4, an architecture for a Next-Generation IDS is presented. An abstract view of the architecture is shown in Figure 2.

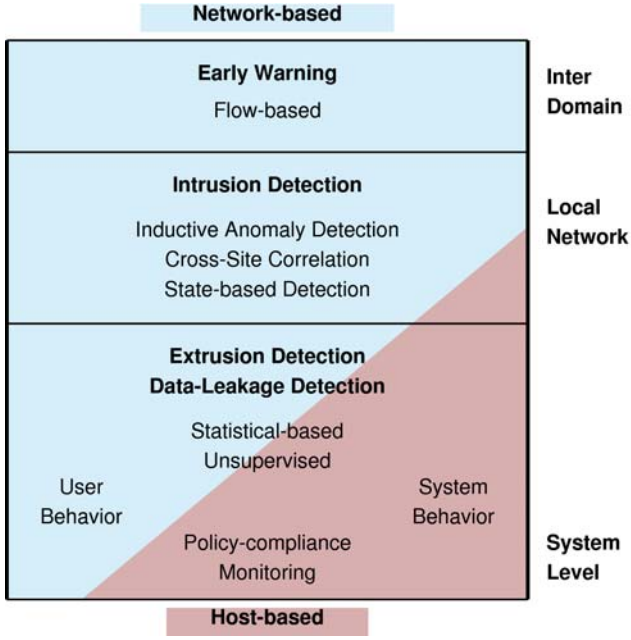

Figure 2. Layers of a Next-Generation IDS.

The system consists of three main parts, Early Warning, Intrusion- and Extrusion Detection. The different parts can be implemented distributed and autonomous. An EWS has to be integrated comprehensive over the Internet. Event correlation, anomaly detection and inter domain cross correlation can be used to detect new threats. This knowledge can then be used to secure other, yet not affected sub-networks in the Internet. The main purpose of the EWS is the detection and prevention of automated and undirected attacks.
The Intrusion Detection is carried out as NIDS. Multiple detection techniques have to be combined: A behavior-based analysis of the network traffic is done to detect known as well as new, yet unknown threats. The needed model has to be built in an unsupervised fashion in such a way, that no endangered learning phase is needed. If the learning phase cannot be eliminated completely, in contrast to most existing systems, malicious instead of benign data can be used (inductive anomaly detection). Cross-site correlation between systems and networks can be used to reduce the false alert rates of the anomaly detection efficiently. Statistical evaluation has to be done to cope with encrypted traffic. Additional, specialized host-based autonomous agents can be used to assist the evaluation. E.g., agents with state-based detection techniques can be used to identify critical states in a SCADA network: The critical states are well-known in industrial systems, therefore an Intrusion Detection can be realized based on a critical state analysis [74].
Extrusion Detection is the last component. It is also integrated into the NIDS, because due to the risk of insiders, manipulation and the administrative outlay with numerous hosts, host-based detection is not enough. Therefore, the user- and system-behavior is monitored by network-based sensors as well as host-based agents.
With respect to the current research and developments, several open issues arise, especially in the area of In- and Extrusion Detection in encrypted environments. Especially the claim of not using payload-related data to be able to cope with encrypted communication, targeted attacks and unknown threats is rarely address in current research (e.g., see [75]).
There are three basic approaches to carry out Intrusion Detection in encrypted communication, namely:

- Protocol-based: Detection of misuse of the encryption protocol
- Intrusive: Modifications of the network infrastructure or the encryption protocol
- Non-Intrusive: Statistical analysis of encrypted traffic

E.g., ProtoMon is a system developed by Joglekar et al. [76] which instruments shared libraries for cryptographic and application level protocols for conducting intrusion detection. Monitoring is integrated into the protocol handling. By that, attacks on the encryption protocol can be detected. Nevertheless, malicious activities hidden inside the encrypted channel could not be detected.

Intrusive techniques are used by Goh et al. They proposed an IDS for encrypted networks which is able to analyze the payload and simultaneously maintaining the confidentiality of the encrypted traffic [77]. The network traffic is replicated and sent to the receiver and also to the Central IDS (CIDS). The protocol is set onto an underlying VPN and adds an additional layer. The system is able to do payload analysis and to keep the confidentiality, but it strongly depends on modifications of the protocols and infrastructure. Also, additional communications protected by e.g. SSH or TLS cannot be analyzed.
Foroushani et al. proposed a system based on the evaluation of the transferred packet sizes and the time intervals between messages [70]. Attacks are detected without decryption by the use of intrusion signatures which are generated from the frequency of accesses and specifications of the TCP traffic. Anyway, because of a high false alarm rate (about 20 % in the best case), the system is not usable for a production environment. The system requires behavior profiles for the target servers and the exchanged information, which are often not available.
Other work addressing IDSs in encrypted environments can be found, but to the best of our knowledge, all of it can be assigned to one of the three categories named before (e.g., see [78] or [79]). Thus, all of these systems are not appropriate for the defined requirements due to the shortcomings already shown.
An important point of all behavior-based systems are the false alert rates. For a comprehensive development of behavior-based techniques in productive environments, false alerts have to be minimized. The idea of a correlation of ingress traffic from different domains is relatively new and shows promising first results. Boggs et al. were able to demonstrate a Proof-of-Concept with pretty small false alert rates [73]. Further investigations are necessary to improve the shown principles and make them usable for the defined requirements.
In the recent area of DLP, most of the proposed systems are host-based and not able to operate only on a network-based installation (e.g., see [80-82]). Extrusion and data leakage detection is a crucial part of a Next-Generation IDS. Therefore, these techniques have to be analyzed regarding the capability to be adapted to network-based systems.

VI. CONCLUSION AND FURTHER WORK

In the paper, an overview of todays most important security threats was given and observable tendencies were shown. State-of-the-Art IDSs, DLPs and EWSs were presented and their shortcomings analyzed. After that, the requirements for a Next-Generation IDS were derived. The papers shows the open issues and wherever available, recent research addressing these topics. The most important and yet unsolved requirements in the area of encryption and behavior-based analysis were lifted out.

ACKNOWLEDGMENT

The authors wish to thank the members of the Chair for Communication Systems and Internet Services at the Universität der Bundeswehr, headed by Prof. Dr. Gabi Dreo Rodosek, for helpful discussions and valuable comments on previous

versions of this paper. The Chair is part of the Munich Network Management Team.

REFERENCES

[1] Anderson, J., *Computer Security Threat Monitoring and Surveillance*, Fort Washington, April 1980

[2] History of Intrusion-Detection Research at SRI's Computer Science Laboratory, `http://www.csl.sri.com/programs/intrusion/history.html`, last seen on January 2011

[3] Sperotto, A., Schaffrath, G., Sadre, R., Morariu, C., Pras, A., Stiller, B., *An Overview of IP Flow-based Intrusion Detection*, IEEE Survey and Tutorials, Third Issue, 2010

[4] Sabahi, F. and Movaghar, A., *Intrusion Detection: A Survey*, 3rd International Conference on Systems and Networks Communications, ICSNC '08, IEEE Computer Society, 2008

[5] *The Jerusalem Virus*, `http://antivirus.about.com/cs/virusencyclopedia/p/jerusalem.htm`, last seen on January 2011

[6] M86 Security, *Security Labs Report*, Jul 2009-Dec 2009 Recap

[7] *Symantec Internet Security Threat Report*, Trends for 2009, Volume XV, April 2010

[8] McHugh, J. and Christie, A. and Allen, J., *Defending Yourself: The Role of Intrusion Detection Systems*, Software, IEEE, Volume 17, Number 5, 2000

[9] *Symantec Report on Attack Kits and Malicious Websites*, 2010

[10] *Ransomware: Extortion via the Internet*, `http://blogs.techrepublic.com/security/?p=2976`, last seen on January 2011

[11] Young, A., Yung, M., *Cryptovirology: Extortion-Based Security Threats and Countermeasures*, Proceedings of the IEEE Symposium on Security and Privacy, 1996

[12] Beckert, Kathrin, *Sicherheitstipp: Wirtschaftspionage per USB-Stick*, FH Gelsenkirchen, `https://www.it-sicherheit.de/ratgeber/it\_sicherheitstipps/tipp/sicherheitstipp-wirtschaftspionage-per-usb-stick/`, last seen on March 26th, 2011

[13] McCormick, M., *Data Theft: A Prototypical Insider Threat*, Advances in Information Security, Volume 39, April 2008

[14] *Ministry of Defence in new data loss scandal*, October 10th, 2008, `http://www.cio.co.uk/news/3225/ministry-of-defence-in-new-data-loss-scandal/`, last seen on January 2011

[15] *Data loss incident affects NASA*, December 10th, 2010, `http://www.backup-technology.com/5451/data-loss-incident-affects-nasa/`, last seen on January 2011

[16] PricewaterhouseCoopers, Martin Luther University Halle-Wittenberg, *Wirtschaftskriminalität 2009*, `http://www.pwc.de/de/risiko-management/assets/Studie-Wirtschaftskriminal-09.pdf`, 2009

[17] Baker, W.H., Hylender, C.D., Valentine, J.A., *2008 Data Breach Investigation Report*, Verizon Business RISK Team, Verizon Business, 2008

[18] Baker, W.H. et al., *2010 Data Breach Investigation Report*, Verizon Business RISK Team, Verizon Business, 2010

[19] *Firesheep Addon for HTTP session hijacking attacks*, `http://codebutler.github.com/firesheep/`, last seen on January 2011

[20] *Firefox extension steals Facebook, Twitter, etc. sessions*, `Firefox extension steals Facebook, Twitter, etc. sessions`, last seen on January 2011

[21] Kaushik, Das, *IPSec & IPv6 - Securing the NextGen Internet*, `http://ipv6.com/articles/security/IPsec.htm`, last seen on March 22th, 2011

[22] Amsterdam Internet Exchange, *sFlow Stats*, http://www.ams-ix.net/sflow-stats/, last seen on March 22th, 2011

[23] The H Security, *Adobe patches 23 holes in Reader and Acrobat*, http://www.h-online.com/security/news/item/Adobe-patches-23-holes-in-Reader-and-Acrobat-1102416.html, last seen on January 2011

[24] The H Security, *SAP introduces a patch day*, http://www.h-online.com/security/news/item/SAP-introduces-a-patch-day-1079976.html, last seen on January 2011

[25] *SourceFire SNORT*, http://www.snort.org/, last seen on January 2011

[26] Ghosh, Anup and Michael, Christoph and Schatz, Michael, *A Real-Time Intrusion Detection System Based on Learning Program Behavior*, Recent Advances in Intrusion Detection, Lecture Notes in Computer Science, LNCS 1907, Springer Berlin / Heidelberg, 2000

[27] *Computer Network Defence Operational Picture (Talisker Radar)* http://www.securitywizardry.com/radar.htm, last seen on January 2011

[28] Lippmann, R., Webster, S., Stetson, D., *The effect of identifying vulnerabilities and patching software on the utility of network intrusion detection*, Proceedings of the 5th international conference on Recent advances in intrusion detection, RAID 2002

[29] The H Security, *Microsoft issues warning about critical IE hole*, http://www.h-online.com/security/news/item/Microsoft-issues-warning-about-critical-IE-hole-1158684.html, last seen on January 2011

[30] Metasploit Framework *ms11_xxx_createsizeddibsection.rb*, https://www.metasploit.com/redmine/projects/framework/repository/revisions/11466/entry/modules/exploits/windows/fileformat/ms11_xxx_createsizeddibsection.rb, last seen on January 2011

[31] Thottan, Marina and Ji, Chuanyi, *Anomaly Detection in IP Networks*, IEEE Transactions on Signal Processing, Volume 51, No. 8, 2003

[32] Bitter, C. and Elizondo, D.A. and Watson, T., *Application of artificial neural networks and related techniques to intrusion detection*, The 2010 International Joint Conference on Neural Networks (IJCNN), IEEE, 2010

[33] Tandon, Gaurav and Chan, Philip and Mitra, Debasis, *MORPHEUS: motif oriented representations to purge hostile events from unlabeled sequences*, Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security, VizSEC/DMSEC '04, ACM, 2004

[34] Li, Yang and Fang, Binxing and Guo, Li and Chen, You, *Network anomaly detection based on TCM-KNN algorithm*, Proceedings of the 2nd ACM symposium on Information, Computer and Communications Security, ASIACCS '07, ACM, 2007

[35] Almgren, Magnus and Jonsson, Erland, *Using Active Learning in Intrusion Detection*, IEEE Computer Security Foundations Workshop, CSFW 04, Volume 17, IEEE, 2004

[36] *Lancope Network Behavior Analysis*, http://www.lancope.com/solutions/network-behavior-analysis.aspx, last seen on January 2011

[37] *FlowMatrix Network Behavior Analysis System*, http://www.akmalabs.com/flowmatrix.php, last seen on January 2011

[38] Debar, Herve and Dacier, Marc and Wespi, Andreas, *A Revised Taxonomy for Intrusion-Detection Systems*, IBM Research, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland, 1999

[39] Bolzoni, D. and Etalle, S., *Approaches in anomaly-based network intrusion detection systems*, Intrusion Detection Systems, Springer, 2008

[40] Winter, Philipp and Hermann, Eckehard and Zeilinger, Markus, *Inductive Intrusion Detection in Flow-Based Network Data Using One-Class Support Vector Machines*, 2011 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS), IEEE, 2011

[41] Liu Hui and Cao Yonghui, *Research Intrusion Detection Techniques from the Perspective of Machine Learning*, 2010 Second International Conference on Multimedia and Information Technology (MMIT), Volume 1, IEEE, 2010

[42] Hu, W. and Hu, W. and Maybank, S., *Adaboost-based algorithm for network intrusion detection*, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Volume 38, Issue 2, IEEE, 2008

[43] Carrascal, Alberto and Couchet, Jorge and Ferreira, Enrique and Manrique, Daniel, *Anomaly Detection using prior knowledge: application to TCP/IP traffic*, Artificial Intelligence in Theory and Practice, IFIP International Federation for Information Processing, Volume 217, Springer, 2006

[44] Casas, Pedro and Mazel, Johan and Owezarski, Philippe, *Steps Towards Autonomous Network Security: Unsupervised Detection of Network Attacks*, 2011 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS), IEEE, 2011

[45] Sourdis, I. and Pnevmatikatos, D.N. and Vassiliadis, S., *Scalable multigigabit pattern matching for packet inspection*, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Volume 16, pages 156 – 166, IEEE, 2008

[46] Kong, Chao and Yang, Bo and Jia, Zhiping and Chen, Zhenxiang, *A Common On-board Hardware Architecture for Intrusion Detection System*, 2009 International Conference on Multimedia Information Networking and Security, IEEE Computer Society, 2009

[47] Korenek, Jan and Kobiersky, Petr, *Intrusion Detection System Intended for Multigigabit Networks*, Design and Diagnostics of Electronic Circuits and Systems, IEEE Computer Society, 2007

[48] Lin, Cheng-Hung and Chang, Shih-Chieh, *Efficient Pattern Matching Algorithm for Memory Architecture*, IEEE Transaction on Very Large Scale Integration (VLSI) Systems, Volume 19, January 2011

[49] Mitra, A., Najjar, W., Bhuyan, L., *Compiling PCRE to FPGA for Accelerating SNORT IDS*, ACM/IEEE Symposium on Architectures for Networking and Communications Systems, 2007

[50] Baker, Zachary K. and Prasanna, Viktor K., *High-throughput Linked-Pattern Matching for Intrusion Detection Systems*, Symposium on Architectures for Networking and Communications Systems, ANCS 05, ACM, 2005

[51] Sourdis, Ioannis and Pnevmatikatos, Dionisios, *Fast, Large-Scale String Match for a 10Gbps FPGA-Based Network Intrusion Detection System*, Field-Programmable Logic and Applications, Springer, 2003

[52] Ming Gao and Kenong Zhang and Jiahua Lu, *Efficient packet matching for gigabit network intrusion detection using TCAMs*, 20th International Conference on Advanced Information Networking and Applications, AINA '06, 2006

[53] Otey, M. and Parthasarathy, S. and Ghoting, A. and Li, G. and Narravula, S. and Panda, D., *Towards nic-based intrusion detection*, Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2003

[54] De Bruijn, W. and Slowinska, A. and Van Reeuwijk, K. and Hruby, T. and Xu, L. and Bos, H., *SafeCard: A Gigabit IPS on the Network Card*, Recent Advances in Intrusion Detection, Lecture Notes in Computer Science, LNCS 4219, Springer, 2006

[55] *IP Flow Information Export (IPFIX) Applicability*, `http://tools.ietf.org/html/rfc5472#section-3.6.2`, last seen on January 2011

[56] *Cisco Systems NetFlow Services Export Version 9*, `http://www.ietf.org/rfc/rfc3954.txt`, last seen on January 2011

[57] *InMon Corporation's sFlow: A Method for Monitoring Traffic in Switched and Routed Networks*, `http://www.ietf.org/rfc/rfc3176.txt`, last seen on January 2011

[58] *Scrutinizer NetFlow & sFlow Analyzer*, `http://www.plixer.com/products/free-netflow.php`, last seen on January 2011

[59] *NetFlow Analyzer*, http://www.manageengine.com/products/netflow/download.html, last seen on January 2011

[60] Lim, Shu Yun and Jones, Andy, *Network Anomaly Detection System: The State of Art of Network Behaviour Analysis*, International Conference on Convergence and Hybrid Information Technology, 2008. ICHIT '08, IEEE Computer Society, 2008

[61] Lawton, G., *New Technology Prevents Data Leakage*, Computer Journal, Volume 41 Issue 9, September 2008, 10.1109/MC.2008.394

[62] *Security-Enhanced Linux*, http://www.nsa.gov/research/selinux/, last seen on January 2011

[63] *Data Loss Prevention Services*, http://us.trendmicro.com/us/products/enterprise/data-loss-prevention/services/, last seen on January 2011

[64] *Fidelis Security Systems appliances and support*, http://www-935.ibm.com/services/us/index.wss/offering/iss/a1031185, last seen on January 2011

[65] *Internet Storm Center* of the SANS (SysAdmin, Audit, Network, Security) Institute, isc.sans.org, last seen on January 2011

[66] *Active Threat Level Analysis System*, Arbor Networks, atlas.arbor.net, last seen on January 2011

[67] Bustamante, F., Choffnes, D., *NEWS plugin for the Vuze BitTorrent Client*, http://www.aqualab.cs.northwestern.edu/projects/NEWS.html, last seen on January 2011

[68] *Early Warning and Intrusion Detection based on Combined AI Methods*, www.fides-security.org, , last seen on January 2011

[69] *Worldwide Observatory of Malicious Behaviors and Attack Threats*, wombat-project.eu, last seen on January 2011

[70] Foroushani, V.A., Adibina, F., Hojati, E., *Intrusion Detection in Encrypted Accesses with SSH Protocol to Network Public Servers*, Proceedings of the International Conference on Computer and Communication Engineering 2008, May 13-15, Kuala Lumpur, Malaysia

[71] Melnikov, N., Schönwälder, J., *Cybermetrics: User Identification Through Network Flow Pattern Analysis*, EMANICS Workshop on NetFlow/IPFIX Usage, Jacobs University Bremen, October 2009

[72] Moraga, C., *Design of Neural Networks*, Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science, Volume 4692/2008

[73] Boggs, N., Hiremagalore, S., Stavrou, A., Stolfo, S., *Experimental Results of Cross-Site Exchange of Web Content Anomaly Detector Alerts*, IEEE Conference on Technologies for Homeland Security, Boston, 2010

[74] Carcano, A. and Coletta, A. and Guglielmi, M. and Masera, M. and Nai Fovino, I. and Trombetta, A., *A Multidimensional Critical State Analysis for Detecting Intrusions in SCADA Systems*, Industrial Informatics, IEEE Transactions on, 2011

[75] Schaffrath, G., *Network Intrusion Detection Systems & Encryption: Friends or Foes?* , Communication Systems Group, University of Zürich, August 2008

[76] Joglekar, S., Tate, S., *ProtoMon: Embedded Monitors for Cryptographic Protocol Intrusion Detection and Prevention*, Journal of Universal Computer Science, Volume 11, 10.3217/jucs-011-01-0083

[77] Goh, V.T., Zimmermann, J., Looi, M. (2010), *Experimenting with an Intrusion Detection System for Encrypted Networks*, Int. J. Business Intelligence and Data Mining, Vol. 5, No. 2, pp. 172-191

[78] Yasinsac, A., Goregaoker, S., *An Intrusion Detection System for Security Protocol Traffic*, Department of Computer Science, Florida State University

[79] Yamada, A., Miyake, Y., Takemori, K., Studer, A., Perrig, A., *Intrusion Detection for Encrypted Web Access*, AINAW 2007, ISBN 0-7695-2847-3

[80] Papadimitriou, P., Garcia-Molina, H., *Data Leakage Detection*, IEEE Transactions on Knowledge and Data Engineering, Volume 23, Number 1, January 2011

[81] Cui, W., Katz, H., Tan, W. *Design and Implementation of an Extrusion-based Break-In Detector for Personal Computers*, Proceedings of the 21st Annual Computer Security Applications Conference (ACSAC 2005), 1063-9527/05

[82] Martignoni, L., Stinson, E., Fredrikson, M., Jha, S., Mitchell, J., *A Layered Architecture for Detecting Malicious Behaviors*, RAID 2008, LNCS 5230

2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

Using a Novel Behavioral Stimuli-Response Framework to Defend against Adversarial Cyberspace Participants

Daniel Bilar
Department of Computer Science
University of New Orleans
New Orleans, LA 70148, USA
dbilar@uno.edu

Brendan Saltaformaggio
Department of Computer Science
University of New Orleans
New Orleans, LA 70148, USA
bsaltafo@uno.edu

Abstract- **Autonomous Baiting, Control and Deception of Adversarial Cyberspace Participants (ABCD-ACP) is an experimental defensive framework against potentially adversarial cyberspace participants, such as malicious software and subversive insiders. By deploying fake targets (called baits/stimuli) onto a virtualized environment, the framework seeks to probabilistically identify suspicious participants through aggregate suspicious behavior, subvert their decision structure and goad them into a position favorable to the defense. Baits include simulating insertion of readable and writable drives with weak or no password, marked doc/pdf/txt/exe/cad/xls/dat files, processes with popular target names and processes that detect thread injections.**

This approach bears some similarities to the concept of subverting an enemy's OODA (Observe, Orient, Decide, and Act) loop, an information warfare strategy which seeks to proactively influence and change enemy behavior. By controlling perception of the environment, this approach similarly seeks to influence adversarial participants' decision complexity, noise levels, effectiveness and ultimately their ability to fulfill their mission. This is a work in progress: The conceptual framework is described, and implemented baits and preliminary empirical results are presented.

The long term project end vision is an autonomic framework playing a repeated, dynamic, imperfect information, non-cooperative stimuli-response game which probabilistically identifies, then impedes, quarantines, subverts, possibly attributes and possibly inoculates against suspected adversarial cyberspace participants.

Keywords: virtualization, malware, dynamic game, stimuli, behavior

This research was funded in part by the US Department of Defense through a DoD IASP Grant (H98230-09-1-0369) which is administered by the US National Security Agency and in part by a grant from the University of New Orleans Office of Sponsored Research.

I. INTRODUCTION

It is written that a person's character may be recognized by how he handles alcohol, his conduct in financial matters and his anger. In other words, behavior shown in certain situations gives insight into character. A cyber-defensive approach in the form of a behavioral stimuli-response framework is presented in this paper. It should be noted that is work in progress.

Why is this needed? It is needed as an addition to Defense-in-Depth. The empirical performance of the first line of defense - anti-viral (AV) byte signature blacklisting - has been steadily declining. Independent laboratory test results over a period of a decade have shown a steady rise in false negative rates, i.e. failing to detect malicious code. The average miss rate of even previously submitted malicious code hovers in the double digits. In 2010, after failing to update static signatures *for just one week*, the best AV tested missed 37%, the worst between 60% and 90% (see TABLE I.).

TABLE I. DETECTION RATE RANGES OF SIXTEEN TO TWENTY POPULAR AV SCANNERS [1]

Report Date	AV Signature Update	Malicious Code Corpus Date	False Negatives (%)
2010/11	Aug. 16th	Aug. 17th-24th	[38-63]
2010/08	Aug. 16th	Aug. 6th	[0.2-19.1]
2010/05	Feb. 10th	Feb 11th-18th	[37-89]
2010/02	Feb. 10th	Feb. 3rd	[0.4-19.2]
2009/11	Aug. 10th	Aug. 11th-17th	[26-68]
2009/08	Aug. 10th	Aug. 10th	[0.2-15.2]
2009/05	Feb. 9th	Feb. 9th -16th	[31-86]
2009/02	Feb. 9th	Feb. 1st	[0.2-15.1]
2008/11	Aug. 4th	Aug. 4th -11th	[29-81]
2008/08	Aug. 4th	Aug. 1st	[0.4-13.5]
2008/05	Feb. 4th	Feb. 5th -12th	[26-94]
2008/02	Feb. 4th	Feb. 2nd	[0.2-12.3]

In addition, recent advances in formal computer virology show that detection of malicious code that poses the most problems (staged downloads and interactive) cannot be accomplished in linear time and enters the realm of exhaustive search space and undecidability. Reference [2] proved that detection of interactive malicious code is at least in complexity class $NP^{((NP^{\wedge}oracle)^{\wedge}(NP^{\wedge}oracle))}$.

This is no accident since the design and implementation of modern malware seeks to specifically undermine the information gain of static signature approaches, in effect presenting the defense with Halting-type problems. The reverse is *not* true: From the point of view of adversarial participants, cyber-targets are pathologically honest and do not systematically confuse adversarial participants with high entropy schemes.

This paper's view is that defenses have to *adopt similar comprehensive dissimulation and deception stances on cyber-targets and embedding environments*. By turning the tables on potentially Adversarial Cyberspace Participants (ACP), their code footprint, decision complexity, noise levels and uncertainty about the 'real' view of the cyber-environment are increased, thereby giving defenses more temporal and spatial leeway. This experimental framework is not meant to substitute for but rather complement traditional blacklisting byte signature based mitigation approaches whose limitations are well-known [3].

II. PRIOR WORK

This paper emphasizes the primacy of ACP control flow subversion through judicious manipulation of the environment's observables. In the realm of best-of-breed static structural signatures, [4] extracted malware family signatures using maximum graph homomorphism between the function callgraph of two executables. The flowgraph structure and its code 'neighborhood' were used to develop an opcode-sequence agnostic graph hash for fast approximate comparison. The intersection of known family members subsequently generated the family superstructure signature. Such a signature extracted automatically from 15 variants of the polymorphic malware family 'Swizzor' was able to recognize 900 additional variants (in a sample of 20,000 unsorted pieces of malware) with no false positives. However, this approach required pertinent structural information to be recovered; a proposition that does not hold with malware that purposely obfuscates its control flow structure.

This paper also posits ACPs (especially malicious software) to be sensitive to real or perceived operating environment changes. For evidence consistent with this assertion, the reader is referred to the 2005 analysis of the Slammer worm, in which complex dependencies between user/kernel processes and threading are described, as well as the 2008/2009 Conficker A worm, which exits upon detection of a Ukrainian keyboard locale [5] [6]. In a comprehensive 2008 empirical study, [7] investigated the environmental awareness of modern malware by measuring the deterrence value of imitating virtual machines and debuggers through light-weight registry key insertions, system call hooking (e.g. *CheckRemoteDebuggerPresent()* set to TRUE) and process generation (e.g. a process named OllyDbg, a popular debugger). Of the 6205 malware samples, about 25% reduced their malicious behavior through these light-weight techniques.

This paper's approach furthermore seeks to draw ACPs into a repeated stimuli-response game with the expectation that its dynamic behavior can be influenced quicker than non-malicious participants. In a comparative analysis of malicious and non-malicious software, [8] showed through statistical static structural analysis that malicious code tended to have a lower basic block count, implying a *simpler decision structure*: less interaction, fewer branches and limited goals compared to

non-malicious software. This suggests that malicious code can be ‘outplayed’ by exploiting this simpler decision structure.

From an implementation point of view, honeypots and honeynets - simulated decoys that detract from 'real' networks, hosts and services – are well known examples of ‘morphing the network’, i.e. changing the perception of the network’s makeup. Reference [9] implemented a highly scalable, parsimonious hybridization of low- and high-interaction honeynets that doubled as a platform for malware collection. He suggested it to be used as part of an automated, next-generation system to stop botnets. Ad-hoc hot patching, as well as randomization techniques (randomized heap/stack/library positioning at compile, link and load times) are incorporated into modern operating systems like Windows Vista/7 [10].

Lastly, probabilistic identification and control of hitherto-unknown/unseen threats serves to enhance situational and behavioral awareness on a host, network and mission level. In this, this project complements other efforts in US military domains: DARPA’s Integrated Battle Command (BAA 05-14) gives decision aids for battle operations, DARPA’s Real-Time Adversarial Intelligence & Decision Making (BAA 04-16) tries to help battlefield commanders compute and counteract threat predictions in tactical operations. Lastly, Israel’s Virtual Battle Management AI - a defense system designed to handle situations that exceed the physiological limits of human command in case of a doomsday strike - mirrors most closely the project end vision [11].

III. DESIGN OF FRAMEWORK

The **Gameboard** consists of a virtualized operating environment (a Windows XP SP2 VM) which is ‘morphed’ by the Defender. Morphing means that from the point of view of Gameboard participants, the environment (or merely its perception by the participants) is altered via stimuli in order to provoke a reaction that could be used for identification. Stimuli (such as a .pst file, a simulated network drive, or a process named iexplore.exe) are introduced to induce potential adversarial participants (both humans and programs) to ‘show their colors’ (see Figure 1).

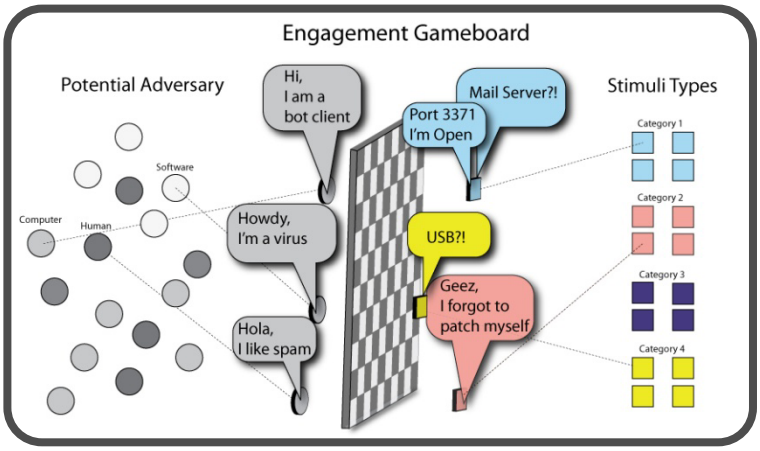

Figure 1: Notional Gameboard illustration. Stimuli (e.g. fake network drives, fake processes with names of popular applications, AutoCad files) are deployed and participants' responses to the baits evaluated.

Conceptually, a repeated, dynamic, imperfect information, non-cooperative stimuli-response game is played on the Gameboard. The players in the game are {Defender} versus {Participants}. All Participants (benign or malicious) are situated within the Gameboard (the VM). The Defender is situated *outside* the Gameboard to hide some of its footprint, but it has the ability to introduce baits/stimuli, change (real or perceived) macroscopic Gameboard parameters, gauge responses and initiate defensive moves.

The game's first goal is to judge whether after several rounds of the stimuli-response game the aggregate evidence warrants classifying a participant's observed behavior as adversarial. The concept of aggregate evidence borrows from Whewell's "Consilience of Induction", in which the convergence of several, ideally independent hypotheses serves to strengthen that conclusion [12]. Upon probabilistic identification, the game's second goal is to engage appropriate defensive measures to impede, quarantine, and subvert the ACP threat.

The working assumptions are as follows:

1. From observations of triggered stimuli and responses, uncertainty anent unknown intent can be reduced. In particular, potential adversarial participants can be probabilistically identified.

2. Defender can control the runtime behavior of ACPs by influencing what Participants perceive within the Gameboard.

A. *Baiting Adversarial Cyberspace Participants*

The repeated stimuli-response game can be conceptually decomposed as follows: A Defender conversation consists of a high level scenario which is either preemptively engaged, chosen by the user, or activated by other defensive systems (such as an NIDS). Conversation examples include "Worm", "Rootkit", "Bot", "Trojan", "Trusted Insider", "Hapless User" and more.

A Defender scenario informs one or more engagement types. Engagement type examples include "Offer spread vectors", "Offer confidentiality vectors", "Offer reconnaissance vectors", "Present weakened defenses", "Change system parameters" and more.

For each engagement type, the Defender autonomously chooses a dynamic engagement strategy. These engagement strategies consist of a game tree aggregate of baits/stimuli, participant responses and defensive responses. It is a dynamic game tree since moves are generated dynamically based on observed responses to previous stimuli.

B. *Controlling and Deceiving Cyberspace Participants*

Collberg's atomic primitives constitute *abstract categories of defenses* and are subsequently used as a blueprint for defensive responses upon probabilistic ACP identification [13]. These primitives are cover, duplicate, split/merge, reorder, map, indirect, mimic, advertise, detect/ response, and dynamic. The framework's adaptation of some of these primitives is given below:

Quarantine [Indirect]: Defender moves ACP to an instrumented but isolated platform in order to learn more about its behavior.

(Self-)termination [Tamperproof]: Defender terminates ACP or induces its self-termination. In addition, the Defender may simulate termination of benign components as a strategic mimetic move (such as unlinking it from the process table).

Scarcity [Mimicry, Tamperproof]: Defender presents the Gameboard in a "critical" or "strained" state in an effort to violate ACP's expected usage scenario (e.g. 99% memory utilization, heavy network congestion, no heap space left) [14].

Subversion [Tamperproof]: Data-taint/poison the input to ACPs in order to create an attribution trail (e.g. email bugs in .pst files). This is especially important for military defense systems, where attackers try to plausibly deny responsibility through one or more levels of indirection.

C. *Composition of Context-Sensitive Interactions*

It is an open research question whether engagement strategies can be derived from first principles (i.e. formal malware models [15]). Similarly, it is not clear a priori which set of defensive responses is best suited for which ACP classes. Empirical sandbox observation of 10,000s of malware samples (exhibiting a wide variety of behaviors) was scheduled, and samples were procured from a friendly malware repository, http://offensivecomputing.net. It turned out that lack of sample metadata (names were hashed) hindered the establishment of 'ground truth' (known identities of the control samples) anent the engagement strategies and the defensive responses. Hence, a systematic evaluation of the dynamic compositional question and concomitant quantitative measurements has not yet been undertaken.

D. *Views of the Gameboard*

Since they are situated within the Gameboard, all Participants have a view of the Gameboard, but not necessarily the same in terms of scope and fidelity. In particular, Participants' views and subsequent behavior are constructed by interacting with the Gameboard (checking if a certain process is running for instance).

Defender's view: All of the Participants' behavior unfolds over time. Some behavior on the Gameboard is benign, while some is potentially adversarial. Some behavior is seen by the Defender via baits that are triggered, while some behavior will not be seen. The Defender engages in conversations with Participants to figure out potential benevolence/malevolence.

Participant's view: The interactions between the Defender (through the Gameboard morphing) and the Participant influence the Participant's perception of the environment and, as posited, subsequent Participant behavior. This behavior may in turn influence the Defender's strategies, and so on, until identification decision thresholds are reached and defensive responses are engaged.

E. *Goals of The Defender*

Drawing from prior experience and input from stakeholders, a list of Defender goals was assembled, in descending order of importance. These goals influence both the nature of the baits/stimuli injected into the Gameboard, the timing of their deployment, as well as defensive responses.

Mission Continuity: Defender should not self-sabotage or sabotage the mission of benign Participants in the Gameboard. The primary goal of any defense is to sustain the mission. Mission continuity constraints include but are not limited to: sustaining mission availability, confidentiality, integrity, and command and control authenticity.

Actionable Information Gain: Defender's responses should be geared towards reducing uncertainty and learning more about potential ACPs. This is in part accomplished by the interactions in the dynamic game. In addition, freezing the Gameboard and migrating the ACP threads into a more highly instrumented environment is being explored.

Defender Stealth: Potentially adversarial participant should remain unaware of Defender's observation and manipulation of ACP's perception of the Gameboard. This is accomplished by positioning the Defender outside of the Gameboard and by randomizing design and implementation aspects of the baits.

Subversion: Defender responds in such a way as to repurpose the adversarial participant for the benefit of the Gameboard's mission. One possibility is supplying the ACPs with specially crafted random input, which has been shown to crash in other contexts between 25%-40% of given applications [17][18].

Participant Attribution: Defender responds in such a way that attribution of an adversarial behavior source is made more likely (e.g. smart watermarking/ poisoning of data).

Inoculation: Defender may be able to synthesize a general modus operandus over observed behavior for the purpose of inoculation: Through judiciously chosen baits the traversal of appropriate control flow paths in the ACP is induced. This is in keeping with the light-weight shutdown results of [7].

IV. IMPLEMENTATION OF THE FRAMEWORK

The Defender needs to influence and control the Gameboard environment in a way that is transparent to the Participants. The VMWare platform was chosen due to its market share and proprietary design. Unlike Bochs or Qemu, VMWare's code is not normally available, forcing manipulation of the VM from the 'outside', with no detectable footprint in the Gameboard besides the baits. Since there are numerous robust VM detection approaches, it is reasonable to assume that Participants can ascertain whether they are running in a virtualized environment [19]. As virtualized execution environment are becoming more commonplace with the push towards large-scale virtualized commercial environments, this is a reasonable extension and benefits the scheme.

A. VMUtils library

VMware's VIX API is used to control the Virtual Machine [20]. Since VIX is still in flux, further modifications led to the development of the library *VMUtils*.

VMUtils wraps a number of VMWare's VIX library functions in order to simplify calls to the VIX API. An example is getting a handle for a VM which previously consisted of multiple lengthy and confusing VIX API calls, all of which had to be

paired with additional error checks. The VMUtils library abstracts bait implementation away from the Defender's engagement strategies and allows for bait design through a mediate layer. This is also useful for a centralized VM administration approach like VMWare Server. Alternatively, a conventional network-based communication API could be substituted in place of VMUtils.

B. *Baits/Stimuli*

Baits/stimuli seek to alter the perception of the operating environment (i.e. 'morph the Gameboard') in order to induce tell-tale behavioral responses from potentially adversarial cyberspace participants. Some changes in the environment are lightweight, sometimes they are entirely simulated:

- Simulating insertion of readable and writable media
- Simulating creation of Network Drives with weak or no password
- Planting marked doc/pdf/txt/exe/cad/xls/dat files
- Planting bank cookies
- Creating fake processes with names of popular AV programs
- Creating processes to detect thread injection
- Navigating a browser to Microsoft Update/AV sites (to see whether access to these sites is blocked)
- Navigating a browser to a bank site to see if participants attempt a XSRF attack
- Navigating a browser to a social network site known to be vulnerable to XSS attacks
- Simulating a particular bot client
- Slowing down or speeding up Gameboard system time

A robust bait portfolio must give quantitative metrics on adversarial participant specificity and sensitivity: Low false positives are desired (i.e. does it flag benevolent participants as adversarial?), as well as low false negatives. This is ongoing empirical work and has not yet been addressed. The baits developed and deployed to date are described below.

1) Dummy Process

A dummy process execution and monitoring bait was implemented first. A well-known ACP tactic is, after infecting a machine, to turn off or uninstall AV software. This bait hence targets the self-defense trait of ACPs by executing a number of bait processes named after popular AV programs and monitors them for execution disruption. Alternatively, it is possible to implement a callback-model for the dummy process baits: the bait program creates new threads for each new bait AV process started within the Gameboard, then makes the thread wait for an exit code from the bait process. The later design was chosen.

A list of common AV process names (e.g. avguard.exe) was compiled into a *config* file, which is read by a baiting program. The baiting program then renamed the dummy process, copied it down to the Gameboard, and executed the bait AV process. By waiting for exit codes from the processes running in the VM, the Defender determined if any (and how many) baits were tripped – in other words, which bait AV programs were terminated.

There are very few legitimate reasons (Force Quit, for instance, being an exception) a non-malicious program would kill a running process of a common AV. Intuitively, this bait has high malicious code specificity. It may have low sensitivity depending on how many ACPs attempt to terminate the dummy processes.

2) Network Shares

Another common ACP tactic is *spreading* via network shares. A mechanism was implemented to mount and remove network shares and monitor them for access; the rationale being that spreading is common for malicious code with network shares representing tempting targets. A Defender directory was mapped to a network drive on the Gameboard. The directory was monitored for changes, immediately alerting the Denfender if an attempt was made to write to the network share.

Since USB keys were used by Conficker and the 2008 Central Command attack for spreading, attempts were undertaken in conjunction with the network shares. It turned out to be harder than anticipated, due to the way VMWare handles USB devices.

3) Data and System Files

A similar mechanism can be used for bait files. As the January 2010 Aurora attack showed, industrial espionage targets the confidentiality of intellectual property, such as AutoCad design files. By data-tainting a seemingly high value file, it is hoped that an attribution trail can be established. Steganographic means may be pursued, but a simpler mechanism was chosen for proof-of-concept.

A data file was created on the host machine containing a bogus .gov or .mil email address (or other attractive metadata), then copied into the Gameboard and monitored for activity. This bait aims to coax out malicious actions of potential ACPs: Some instances of malicious code will search a filesystem looking for anything that looks like email addresses, accounting spreadsheets, Outlook .pst files or other data files. Using the same monitoring program from the network share bait, this bait would be tripped on file access or, at a later point in time, by bogus email usage. Although this bait is straightforward to implement, it may have lower specificity due to installed indexers like Google Desktop. This mechanism may also be applied to sensitive system files, complementing Windows File Protection.

4) AV Sites

Editing the Windows Hosts file is a way that malicious code will attempt to block web access to AV websites. This is an example of an ACP's self-defense trait with high specificity.

A bait program was written to test connectivity to many known AV websites. The bait program read URLs from a *config* file and sent http requests to the web-server from within the Gameboard. The first request was sent to the URL; then, using an external DNS server, the same request was sent to the corresponding IP address. Return codes were then compared to determine if malicious code had tampered with web requests. If no determination could be made, a HTTP request was sent by the Defender from the outside and used as a control to compare the previous samples taken from within. Determining whether malicious code is interrupting connections to AV servers constitutes a highly specific indicator of malicious behavior. In 2010, a similar method was used for the Conficker Eye Chart test to test for Conficker infection.

5) User Activities

Another scheme is to simulate normal user behavior to coax out malicious ACP action. Any form of day-to-day user activity might constitute a trigger for malicious code. Such activities include, but are not limited to, checking email, program execution, online banking, or social networking. These activities are simulated and monitored for interruption or abnormal execution.

As a proof of concept, Visual Basic (VB) scripts were used because of the tight integration with Windows and the MSDN references [21][22]. These scripts were deployed onto the Gameboard. Although this needs to be verified empirically, the script's execution isn't likely to be detectable with an acceptable false positive rate by malicious software because MS Windows' handling of VB scripting through wscript.exe. The observable change the ACP sees is wscript.exe running, but there is no straightforward way to tell what it is doing or that it is a Defender's bait script. An example is given in Figure 2, a Yahoo login script controlling a Firefox browser.

```
Set oShell = WScript.CreateObject("WScript.Shell")          ' Create a Shell object.
Dim cmd
cmd = Chr(34) & "C:\Program Files\Mozilla Firefox\firefox.exe" & Chr(34)
oShell.run cmd,0,True  ' Open Firefox
WScript.Sleep 1000
oShell.sendKeys "%d"                    ' Firefox keyboard shortcut for Select Location Bar
oShell.sendKeys "www.mail.yahoo.com"
oShell.sendKeys "{ENTER}"
WScript.Sleep 10000
oShell.sendKeys "emailAccount"     'User name
oShell.sendKeys "{TAB}"
oShell.sendKeys "letmein"               'Password
oShell.sendKeys "{ENTER}"
WScript.Sleep 1000
oShell.sendKeys "^t"            'Firefox keyboard shortcut for New Tab
oShell.sendKeys "%d"          'Firefox keyboard shortcut for Select Location Bar
oShell.sendKeys "www.evil.com"
oShell.sendKeys "{ENTER}"
```

Figure 2: Navigating Firefox to Yahoo.com with wscript.exe

This example shows logging into a Yahoo mail account, opening a new tab, and navigating to a different website. Should an XSS injection be detected (in conjunction with an open source tool, XSSer), the Defender is notified. This feature is being validated further.

6) Thread Injection

Windows' *CreateRemoteThread(..)* serves as a prevalent exploit vector for malicious code. This function, exposed by Windows executables, enables the injection of an arbitrary thread inside the memory space of other processes [23].

Similar to the Dummy Process bait, a process is deployed to run within the Gameboard, continuously querying its number of threads. Once the bait process detects additional threads, it reports back to its monitor outside the Gameboard and terminates. Detection of remote thread injection was chosen under the assumption that it represents both a highly specific and sensitive trigger of malicious activity in non-debugging environments.

7) Macro-Enviromental Triggers

Under development are so-called *macro-environmental triggers*, such as controlling tick time within the Gameboard. By speeding up or slowing down tick time, time-dependent actions could be triggered, thus allowing for more inference anent ACPs' decision structure, patterns and/or movements. Macro-environmental triggers are by their very nature not highly specific: sudden loss of resources such as RAM/HDD shortages and network congestion have been shown in other contexts to crash programs in unexpected ways [14].

C. *Defender*

As noted, the Defender schedules, organizes, and monitors the baits, as well as coordinates defensive responses. It also keeps track of information about the Gameboard, such as logins and file paths.

The Defender loads information about the baits it intends to run from a *config* file. These baits are then deployed in an order, timing and frequency determined by a dynamic engagement strategy. The baits write information back to the controller via a pipe: millisecond timing analysis, bait trip counts, and errors are subsequently stored in a database. This aggregate evidence is used to weigh different hypotheses (using a Bayesian log likelihood model selection approach [16]) anent the observed behavior and formulate dynamic engagement strategies. This is still under development; in the proof-of-concept prototype, only static strategies have been implemented so far.

V. PRELIMINARY EMPIRICAL VALIDATION

Implemented baits are summarized in TABLE II. The rightmost column lists malicious code examples that informed the design of the baits.

TABLE II. IMPLEMENTED BAITS AND MALWARE TRIGGERS

Bait Name	Bait Action	Malware example
Dummy processes	Inject false antivirus programs into the OS process list and monitor for halt in execution	Conficker [24] (kills AV processes), Bugbear [25] (shuts down various AV processes), Vundo[26] (disables Norton AV)
Network Shares	Mounts and removes network shares on the client then monitors them on the server's side for activity	MyWife.d [27] (attempt to delete System files on shared network drives), Lovgate [28] (copies itself to all network drives on an infected computer), Conficker (infects all registered drives)
Files	Monitors system critical or bait files on the client for activity	Mydoom.b [29] (alters the host file to block web traffic), MyWife.d (deletes AV and system programs), Waledac.a [30] (scans local drives for email addresses)
User Action	Executes "normal" user behavior on the client system and monitors for unusual execution	Mydoom.b (diverts internet traffic, thus altering what is expected to appear), Vundo (consumes system resources and slows or impedes program execution)
Thread Injection	Continually queries its number of threads for any changes from the expected number	Poisonivy [31] (injects code into processes such as 'explorer.exe' or 'msnmsgr.exe'), Pandex [32] (seeks 'iexplore.exe' program to inject its code)

The Win32 MyDoom.b email worm was used to generate the following time line points: t_{0a} (bait setup), t_{0b} (bait deployed and ready to be triggered), t_1 (malicious code is executed), t_2 (bait is triggered), and t_3 (bait is recalled/terminated) as described in Table III.

TABLE III. MYDOOM.B TIMING RESULTS (AVERAGE IN SECONDS)

Bait	t_{0a}	t_{0b}	t_1	t_2	t_3
Files (watching critical directories)	67	68	68	69	70
User action (checking AV websites)	70	71	68	73-103	103

These preliminary timing experiments are consistent with the second assumption stated in the beginning of Section III: ACP runtime behavior can be influenced by

Gameboard perception. Many of our samples actually failed to run within the Gameboard. Upon closer inspection, it seems that the virtualized environment provides a certain amount of protection in itself; malicious software often checks whether it is running in a debugging and/or virtualized environment and subsequently does not exhibit malicious behavior [33].

VI. FEASIBILITY AND FUTURE WORK

It should be clear from the exposition that this experimental framework is merely at an early proof-of-concept stage. Whatever research direction is charted, quantification of metrics and empirical validation are to be addressed since they represent methodological lacunae in the literature. A meta-survey of ninety security papers between 1981 and 2008 showed that quantified security was a weak hypothesis because of lack of validation and comparison against empirical data [34]. Bearing this in mind, future research must additionally tackle the following issues:

Behavior inferred by the stimuli-response framework needs to be modeled. Leveraging previous behavioral ontology work [35] and following Shannon's terminology for Markovian models, a mechanism was recently proposed to extract and characterize cyber-behavioral traits of humans for classification, prediction and change detection purpose. That framework introduced the notion of 0^{th} (atomic elements), 1^{st} (atomic + frequencies + context), and 2^{nd} order (probabilities of sequence of activities + context) behaviors. The approach has been evaluated in domains such as military targeting, stress monitoring, and insider threat detection with encouraging results [36].

From a game-theoretic perspective, the game may be played with *obscuring* participants. Obscuring participants may be able and willing to play sub-optimally (not take baits for example) to thwart behavioral estimates. In the context of cyber adversaries, maximum-entropy and hidden Markov model methods have been used to estimate subgame probabilities (i.e. the proportion of time spent in malicious and benign subgames). This approach may be extended to obfuscating adversaries, who attempt to hide their subgame probabilities [37].

Lastly, in order to transition the framework to production systems, the performance and stability challenges of scaling to 100,000s of virtualized hosts on infrastructure clouds will have to be kept in mind at design time [38].

As noted, the project end vision is an autonomic framework playing a repeated, dynamic, imperfect information, non-cooperative stimuli-response game which probabilistically identifies, then impedes, quarantines, subverts, possibly attributes and possibly inoculates against suspected adversarial cyberspace participants. Speculatively, an autonomous defense 'alter ego' for human decision makers is envisioned which, when coupled with physiological sensors, remains poised to take over when human judgment is deemed to be too affected by emotions and/or

information overload. As far-fetched as this may sound in 2011, skeptical readers are invited to peruse the US Air Force Chief Scientist's vision for 2010-2030 [39].

ACKNOWLEDGEMENTS

We thank the anonymous reviewers at the National Security Agency and at the ICCC3 conference for their helpful suggestions and comments.

REFERENCES

[1] A. Clementi, “Anti-Virus Comparatives,” http://av-comparatives.org, Feb. 2008 - Dec. 2010.

[2] G. Jacob, and E. Filiol, “Malware as Interaction Machines,” J. Comp. Vir. 4:3, 2008, pp. 235-250

[3] M. Locasto, Y. Song, and S. Stolfo, “On the infeasibility of modelling polymorphic shellcode,” in ACM CCS, 2007, pp. 541–551

[4] T. Dullien, and E. Carrera, and S. Eppler, and S. Porst, “Automated Attacker Correlation for Malicious Code,” in Proceedings of the NATO IST Symposium (Tallinn, Estonia), November 2010, pp. 26.1-26.10

[5] J. Crandall, Z. Su, S. Wu, and F. Chong, “On deriving unknown vulnerabilities from zero-day polymorphic and metamorphic worm exploits,” in Proceedings of the 12th ACM CCS, pp.235-248, 2005.

[6] P. Porras, and H. Saidi and V. Yegneswaran,” An Analysis of Conficker”, SRI International Technical Report, March 2009.

[7] X. Chen, “Towards an understanding of anti-virtualization and anti-debugging behavior in modern malware”, ICDSN Proceedings, pp. 177-186, 2008.

[8] D. Bilar, “On Callgraphs and Generative Mechanisms,” in J. Comp. Vir. 3:4, 2007, pp. 285-297

[9] D. Zamboni et al., “The Nepenthes Platform: An Efficient Approach to Collect Malware,” in LNCS 4219, Berlin: Springer, 2006, pp. 165-184.

[10] M. Conover, “Assessment of Windows Vista Kernel-Mode Security”, Symantec Advanced Threat Research, 2006.

[11] N. Shachtman, " Israel Eyes Thinking Machines to Fight ‘Doomsday’ Missile Strikes," Wired Danger Room, http://www.wired.com/dangerroom/2008/01/israel-thinking/, January 2008

[12] L. Snyder, “'The whole box of tools': William Whewell and the logic of induction,” in Handbook of the History of Logic - British Logic in the 19th Century, vol. 4, 2008, pp. 163-228

[13] C. Collberg. “Surreptitious Software: Models from Biology and History,” Computer Network Security Series, Berlin: Springer, 2007, pp. 1-21

[14] H. Thompson, J. Whittaker, and F. Mottay, “Software Security Vulnerability Testing in Hostile Environments.“ in Proceedings of the ACM Symposium on Applied Computing, 2002, pp. 260-264

[15] S. Kramer and J. Bradfield, “A General Definition of Malware,” J. Comp. Vir 6:2, 2010, pp. 105-114

[16] R. Kass and L. Wasserman, “A Reference Bayesian Test for Nested Hypotheses and its Relationship to the Schwarz Criterion,” in Journal of the American Statistical Association 90:341, 1995, pp.928-934

[17] B. P. Miller, G. Cooksey, and F. Moore, “An Empirical Study of the Robustness of MacOS Applications Using Random Testing,” in Proceedings of the 1st International Workshop on Random Testing, 2006, pp. 46-54

[18] B. P. Miller, L. Fredriksen, and B. So., “An Empirical Study of the Reliability of Unix Utilities,” in CACM 33:12, 1990, pp. 32-44

183

[19] P. Ferrie, “Attacks on Virtual Machine Emulators“, Symantec Advanced Threat Research, 2007

[20] VMWare, “VIX API 1.10.2 Documentation”, http://www.vmware.com/support/developer/vix-api/, October 2010

[21] M. Russinovich and D. Solomon, “Microsoft Windows Internals,” 5th ed., Microsoft Press, June 2009.

[22] Microsoft, “VBScript Language Reference”, http://msdn.microsoft.com/en-us/library/d1wf56tt%28v=VS.85%29.aspx, 2011

[23] B. Blunden, “The Rootkit Arsenal: Escape and Evasion in the Dark Corners of the System,” Plano, TX: Wordware Publishing, 2009, pp.245-265

[24] V. Tiu, "Virus Analysis - Confounded Conficker," Virus Bulletin, pp. 7-11, March 2009

[25] F-Secure, “Virus Encyclopedia – Worm:W/32Bugbear”, http://www.f-secure.com/v-descs/tanatos.shtml, 2009

[26] Symantec Corporation, “Security Response – Trojan.Vundo”, http://www.symantec.com/security_response/writeup.jsp?docid=2004-112111-3912-99, 2011

[27] McAfee Inc., “Virus Profile: W32/MyWife.d@MM!M24”, http://home.mcafee.com/VirusInfo/VirusProfile.aspx?key=138027, 2010.

[28] Sophos Ltd., “Sophos Security Analyses – W32/Lovgate-E Win32 worm”, http://www.sophos.com/security/analyses/viruses-and-spyware/w32lovgatee.html, 2011

[29] F-Secure, “Virus Encyclopedia – Email-Worm:W/32Mydoom.B”, http://www.f-secure.com/v-descs/mydoom_b.shtml, 2009

[30] F-Secure, “Virus Encyclopedia – Email-Worm:W/32Waledac.A”, http://www.f-secure.com/v-descs/email-worm_w32_waledac_a.shtml, 2009.

[31] D. Elser, "Metafile Art Class," in Virus Bulletin, June 2008, pp. 4-7

[32] C. Prakash and A. Thomas, "Malware Analysis – Pandex: The Botnet That Could," Virus Bulletin, pp. 4-8, March 2008.

[33] T. Raffetseder, C. Kruegel, and E. Kirda, “Detecting Systems Emulators,” in LNCS 4779, Berlin:Springer, 2007, pp.1-18

[34] V. Verendel, “Quantified security is a weak hypothesis”, in Proceedings of the NSPW, 2009, pp. 37-50

[35] N. Sandell, R. Savell, D. Twardowski, and G. Cybenko, “HBML: A Representation Language for Quantitative Behavioral Modeling in the Human Terrain,” in Social Computing and Behaviorial Modeling, New York: Springer, 2009, pp. 180-190

[36] D. Robinson, “Cyber-Based Behavioral Modeling”, PhD Thesis, Dartmouth College (Thayer School of Engineering), July 2010

[37] J. T. House and G. Cybenko, "Hypergame Theory applied to Cyber Attack and Defense," Proc. SPIE, vol. 7666, 2010

[38] E. Kotsovinos, “Virtualization: Blessing or Curse?, ” in CACM 54:1, pp. 61-65, January 2011

[39] W. Dahms, "Technology Horizons: A Vision for Air Force Science & Technology During 2010-2030," Technical Report, USAF Science and Technology, http://www.af.mil/information/technologyhorizons.asp, May 2010

Author Index