Technology Innovation Management Review

July 2013

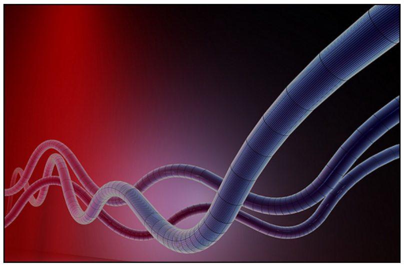

Image licensed under CC BY the M4D Group

Cybersecurity

Welcome to the July 2013 issue of the *Technology Innovation Management Review*. This month's editorial theme is Cybersecurity. We welcome your comments on the articles in this issue as well as suggestions for future article topics and issue themes.

Carleton University

www.timreview.ca

Publisher

The *Technology Innovation Management Review* is a monthly publication of the Talent First Network.

ISSN

1927-0321

www.timreview.ca

Overview

The *Technology Innovation Management Review* (TIM Review) provides insights about the issues and emerging trends relevant to launching and growing technology businesses. The TIM Review focuses on the theories, strategies, and tools that help small and large technology companies succeed.

Our readers are looking for practical ideas they can apply within their own organizations. The TIM Review brings together diverse viewpoints – from academics, entrepreneurs, companies of all sizes, the public sector, the community sector, and others – to bridge the gap between theory and practice. In particular, we focus on the topics of technology and global entrepreneurship in small and large companies.

We welcome input from readers into upcoming themes. Please visit timreview.ca to suggest themes and nominate authors and guest editors.

Contribute

Contribute to the TIM Review in the following ways:

- Read and comment on past articles and blog posts.
- Review the upcoming themes and tell us what topics you would like to see covered.
- Write an article for a future issue; see the author guidelines and editorial process for details.
- Recommend colleagues as authors or guest editors.
- Give feedback on the website or any other aspect of this publication.
- Sponsor or advertise in the TIM Review.
- Tell a friend or colleague about the TIM Review.

Please contact the Editor if you have any questions or comments: timreview.ca/contact

The PDF version is created with Scribus, an open source desktop publishing program.

Editorial: Cybersecurity

Chris McPhee, Editor-in-Chief
Tony Bailetti, Guest Editor

From the Editor-in-Chief

Welcome to the July 2013 issue of the *Technology Innovation Management Review*. This is the first of two issues covering the editorial theme of Cybersecurity, and I am pleased to introduce our guest editor, **Tony Bailetti**, Director of Carleton University's Technology Innovation Management program (TIM; carleton.ca/tim) in Ottawa, Canada.

In addition to six articles and a Q&A on cybersecurity, the July issue also includes a report on a recent TIM Lecture by **Mika Westerlund**, Assistant Professor of Carleton University's Sprott School of Business. In his lecture titled "Green Business Models to Change the World", he presented an array of emerging business models as well as recent research and trends relating to sustainability and green innovation.

In September and October, we will present two issues on Managing Innovation for Tangible Performance, for which the guest editor is **Sorin Cohn**, President of BD *Cohn*sulting Inc. Dr. Cohn also presented the April TIM Lecture on "Enhancing Competitive Position Through Innovation Beyond R&D" (timreview.ca/article/686).

We hope you enjoy this issue of the TIM Review and will share your comments online. Please contact us (timreview.ca/contact) with article topics and submissions, suggestions for future themes, and any other feedback.

Chris McPhee
Editor-in-Chief

From the Guest Editor

It is my pleasure to be the guest editor for the July and August issues of the *TIM Review*. These two issues mark the first milestone of a nationwide effort to make Canada a leader in cybersecurity. This effort sets a new direction for addressing cybersecurity and will be described in the next issue of the journal.

Cyberspace has contributed positively to the world’s economic, political, and social development. However, the integrity of cyberspace is being threatened worldwide. Cyberattacks have become common occurrences and often disrupt existing economic, legal, political, and social agreements. These attacks use well-researched software designed to defeat or bypass security systems, are criminally or politically motivated, and are executed by highly determined, skilled, and well-funded individuals and organizations. Cyberattacks include stealing intellectual property, disrupting national infrastructure, confiscating online bank accounts, creating and distributing viruses, posting confidential information, and encrypting systems to demand ransom.

The July and August issues of the *TIM Review* provide articles that contribute practical experience and academic knowledge that can help Canadians and their allies around the world to benefit from a secure cyberspace. These articles examine the challenges we all face as well as the research, development, entrepreneurial, commercial, and social opportunities that these challenges open up.

Given the interdisciplinary nature of the challenges, these two issues of the journal are the result of close industry, university, and government collaboration. Twelve professionals contributed six articles and a Q&A to the July issue. Five of these authors work in industry, four in universities, and three in government.

Editorial: Cybersecurity

Chris McPhee and Tony Bailetti

Benoit Dupont is the Canada Research Chair in Security and Technology at the Université de Montréal. In this article, he reviews nine socio-technical trends that are likely to shape the cybersecurity environment over the next decade. He examines six cybersecurity implications of these trends and identifies changes in regulations that could help address future cybersecurity issues.

Dan Craigen and **D'Arcy Walsh** are Science Advisors at the Communications Security Establishment Canada, and **David Whyte** is Technical Director for the Cyber Defence Branch at the Communications Security Establishment Canada. These authors outline the elements and conditions required to establish a secure, stable, and resilient information technology infrastructure and formulate a set of principles for the cybersecurity research program to support Canada’s Cybersecurity Strategy.

David Archer is a Research Program Lead at Galois, Inc. and **Adam Wick** directs the Systems and Networking Group at Galois, Inc. Their article discusses an approach that would allow critical information about potential threats to be shared rapidly enough to facilitate a recipient’s timely and effective response. Such relevant sharing of information seldom occurs using existing approaches.

Arthur Low is the founder and Chief Executive Officer of Crack Semiconductor, and **Steven Muegge** is an Assistant Professor at Carleton University's Sprott School of Business. These authors suggest that small, innovative suppliers of network security processors and high-performance security applications should launch and grow a business ecosystem. Organizations that are part of the ecosystem can innovate using a platform of reconfigurable and extensible network security processor technology.

Dan Craigen and **D'Arcy Walsh** are Science Advisors at the Communications Security Establishment Canada, and **Drew Vandeth** is a Senior Researcher at IBM Systems Research and Senior Research Strategist for the National Security Community. Their article describes an approach and operational issues around managing a research and experimental development program that is both adaptive to continuously evolving cybersecurity issues, as well as compatible with international standards published by the Organization for Economic Cooperation and Development and the Treasury Board of Canada Secretariat.

Xinxin Fan and **Guang Gong** are from the Department of Electrical and Computer Engineering at the University of Waterloo. They examine the communication security aspects of a smart-grid metering and control system from the perspective of cryptographic techniques, and they discuss different mechanisms to enhance the cybersecurity of the next-generation power systems.

Sherif Koussa is the founder and Principal of Software Secured. He answers the question "Should startups care about application security?". He argues that executives of successful startups recognize the value of security as a market differentiator and incorporate security in their software from the start to reduce costs.

The integrity of cyberspace is in jeopardy, and we face challenges that require interdisciplinary solutions. We believe that the July and August issues of the *TIM Review* will accelerate industry, government, universities, not-for-profits, and individuals to work together in ensuring that Canadians and their allies benefit from a secure cyberspace. We encourage you, your colleagues, and your organizations to act decisively to make Canada a leader in cybersecurity worldwide and improve the security of cyberspace.

Tony Bailetti
Guest Editor

Editorial: Cybersecurity

Chris McPhee and Tony Bailetti

About the Editors

Chris McPhee is Editor-in-Chief of the *Technology Innovation Management Review*. Chris holds an MASc degree in Technology Innovation Management from Carleton University in Ottawa and BScH and MSc degrees in Biology from Queen's University in Kingston. He has over 15 years of management, design, and content-development experience in Canada and Scotland, primarily in the science, health, and education sectors. As an advisor and editor, he helps entrepreneurs, executives, and researchers develop and express their ideas.

Tony Bailetti is an Associate Professor in the Sprott School of Business and the Department of Systems and Computer Engineering at Carleton University, Ottawa, Canada. Professor Bailetti is the Director of Carleton University's Technology Innovation Management (TIM) program. His research, teaching, and community contributions support technology entrepreneurship, regional economic development, and international co-innovation.

Citation: McPhee, C. and T. Bailetti. 2013. Editorial: Cybersecurity. *Technology Innovation Management Review*. July 2013: 3–5.

Keywords: cybersecurity, cyberthreats, cyberattacks, research, information technology, network security, Canada

Cybersecurity Futures: How Can We Regulate Emergent Risks?

Benoit Dupont

> *“When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.”*
>
> Arthur C Clarke (1917–2008)
> Science-fiction writer, futurist, and inventor

This article reviews nine socio-technical trends that are likely to shape the cybersecurity environment over the next decade. These nine trends have reached various levels of maturity, and some – such as quantum computing – are still theoretically contentious. These trends are: cloud computing; big data; the Internet of Things; the mobile Internet; brain–computer interfaces; near field communication payment systems; mobile robots; quantum computing; and the militarization of the Internet.

What these nine trends have in common is that they will be instrumental in generating new opportunities for offending, which will result from an exponential increase in the quantity of data, number of connection points to the Internet, and velocity of data flows that irrigate the digital ecosystem. As a result, more opportunities for malicious exploitation will be available to attackers, “security by design” will be harder to achieve in such a fluid and dynamic environment, and the performance of control mechanisms is likely to erode significantly.

Technical solutions to address these challenges are already being developed by computer scientists. This article focuses on a different and complementary approach, finding inspiration in the work of regulatory scholars who have framed promising theories such as regulatory pluralism and responsive regulation to explore options for the necessary institutional adaptation to these future changes.

Introduction

The current threat landscape that characterizes computer networks and critical infrastructures is already so saturated with complex risks that it seems futile to extrapolate what the future of cybersecurity will look like 10 years from now. Indeed, Napoleon Bonaparte once said that “simpletons talk of the past, wise men of the present, and fools of the future” (tinyurl.com/7yhoexj). However, a number of information technologies have disrupted existing economical, social, political, and legal arrangements, and it is likely that similar transformations will repeat themselves at regular intervals. The term “disruptive technology” was first used by Clayton Christensen (1997; tinyurl.com/7onvohk) to analyze innovations that do not simply improve the performance of existing technologies (these innovations are called sustaining technologies), but that instead define entirely new products or services to meet unsatisfied needs, and consequently make a lasting change in the technological landscape into which they fit. However, criminals are also very ingenious innovators who take advantage of disruptive technologies to open "breaches", which can be defined as "sudden new opportunities for offending that opened as a result of changes in the technological or social environment" (Killias, 2006; tinyurl.com/m6qmdz5). These breaches are often the result of a defective legal or regulatory coverage and provoke rapid increases in offences, before the breach is closed and offenders move on to the next opportunity. In such a rapidly evolving context, it therefore becomes crucial to anticipate what breaches are likely to open as the result of technological innovations, so that policies and regulations can be developed proactively in order to minimize their impact on Internet users.

Current cybersecurity discourses focus on national security threats such as destructive cyberattacks against critical infrastructures or cyberspying campaigns targeting valuable intellectual property and sensitive strategic information (Brito and Watkins, 2011: tinyurl.com/mtmv7xy; ONCE, 2011: tinyurl.com/638opk9; Gendron and Rudner, 2012: tinyurl.com/lbn2yxm). However, more mundane cybercriminal risks receive considerably less attention and investments from governments, despite the fact that they already affect a much larger share of the population than their national-security counterparts. According to recent Canadian victimization statistics, cyberfrauds represent roughly one-third of all property crimes and significantly outnumbered car thefts, burglaries, and vandalism incidents in 2009 (Perreault and Brennan, 2010: tinyurl.com/lnhlg4a; Perreault, 2011: tinyurl.com/mwae6m6), in line with similar patterns observed in the UK (Anderson et al., 2012; tinyurl.com/csnqtkr). Yet, the majority of police organizations remain under-resourced to address this issue, policy makers are still in the process of developing effective cybercrime control mechanisms, and many private actors keep on marketing equipment, applications, and services whose security remain problematic.

This article sketches the contours of the cybersecurity challenges that are likely to emerge over the next decade and to analyze their security and regulatory implications, so that more effective systems can be designed to monitor and close breaches. In the first section, I introduce the nine disruptive technological trends that forecasters predict will most radically alter the Internet ecosystem over the next 10 years. In the second section, I examine the six cybersecurity implications of these trends and discuss potential breaches that could open if the *status quo* is maintained. Finally, in the third section, I consider what regulatory adaptations could deal more effectively with future cybersecurity problems.

Nine Disruptive Socio-technical Trends

The nine socio-technical trends that are discussed below and are most likely to have an impact on the cybersecurity environment over the next decade were identified and described by the author in a report commissioned by Public Safety Canada Cybersecurity's Directorate and are available online (Dupont, 2012; tinyurl.com/kqqd39f). Because this list includes trends that have reached various stages of maturity, there is unfortunately a strong bias toward technologies that are already commercially available or are reaching the "peak of inflated expectations" in Gartner's "Hype cycle" (Fenn, 2010; tinyurl.com/msw6vn2).

1. Cloud computing
The consulting firm IDC estimates that, in 2020, one-third of computer data will be stored in or will transit through systems administered in the cloud, and that the explosion of this market could generate revenues in excess of one trillion dollars by 2014 (Gantz and Reinsel, 2010: tinyurl.com/m8curcy; Nash, 2011: tinyurl.com/k3egchu). The unparalleled flexibility of cloud computing that promises reduced costs to companies that use it make it an irresistible proposition, particularly in these turbulent financial times (IBM, 2011; tinyurl.com/l7o23cb), and even individuals become avid consumers of cloud services such as Dropbox or Netflix.

2. Big data
The term big data reflects the appearance in recent years of datasets containing gigantic volumes of unstructured or disparate information. The units of measurement used to describe these volumes of data are no longer the gigabyte or the terabyte, but the peta , exa-, or even zettabyte (10^{21} bytes). IDC estimates that, in 2011, the worldwide quantity of information created and exchanged on digital media (the digital universe) was approximately 1.8 zettabytes, and that it would be multiplied by 20 by 2020 to reach 38 zettabytes (Gantz and Reinsel, 2011; tinyurl.com/3f56u9t). The volume and diversity of the data processed prevent traditional analysis techniques from being used, and specialized solutions that are based on cutting-edge computer tools and statistics (such as Hadoop MapReduce programming [tinyurl.com/qqjot] and R language [tinyurl.com/yp9y64] for statistical analyses and visualization) are deployed on infrastructures specially designed for such uses.

3. The Internet of Things
This term refers to the growing interaction between the physical and digital worlds through sensors and data-

capture devices integrated into the objects around us (from cars to pacemakers to refrigerators to smart meters). These objects gain the ability to communicate wirelessly with computer networks through the Internet. The massive flow of data produced by these objects allows for their operations and the environments in which they operate to be more effectively monitored and managed (Chui et al., 2010; tinyurl.com/mqa7942). There are already more objects than computers connected to the Internet (Fenn and LeHong, 2011; tinyurl.com/7a577pl), and Cisco predicts that over 50 billion objects will be connected to the Internet by 2020 (Evans, 2011; tinyurl.com/88uhsx3).

4. Mobile Internet
The concept of mobile Internet or mobile computing designates all technologies that provide full or partial access to the Internet using mobile devices such as smartphones or tablets. In 2012, worldwide sales of mobile Internet devices reached 850 million units, whereas desktop and laptop PCs barely moved 350 million units. The growth rate for tablets and smartphones is evaluated at 174% and 110% respectively over the next four years (IDC, 2013; tinyurl.com/ck2aoxj).

5. Brain–computer interfaces
Brain–computer interfaces are technologies used to directly connect external computer devices to the human brain. These devices allow individuals to interact with computers by thought. These technologies are currently used in medicine to compensate, assist, or augment the cognitive and motor functions of individuals with physical or psychological disabilities. These previously costly technologies that were restricted to the world of research are appearing in consumer electronics and will gradually replace the keyboard and mouse as humans' preferred ways to interact with machines (Yuan and Barker, 2011; tinyurl.com/mkgxm4s). Significant advances have been made in this field, and for the past few months Emotiv (emotiv.com) has been marketing a $300 wireless neuro-headset to capture and process brain signals.

6. Near field communication (NFC)
This is a form of payment that uses various wireless communication technologies related to radio-frequency identification (RFID; tinyurl.com/82u9a) chips to facilitate financial transactions at points of sale. This technology is primarily installed on payment cards and on mobile phones, which can carry out a transaction if placed a few centimetres from a properly-equipped receiver. This technology considerably accelerates the point-of-sale process (Tata, 2011; tinyurl.com/nywjbqx) and is intended to compete directly with traditional payment methods such as cash or credit cards (Ondrus and Pigneur, 2009; tinyurl.com/la3xq9b).

7. Mobile robots
Multi-jointed mechanical systems that are able to travel autonomously or semi-autonomously and that have the ability to influence their immediate environment are known as mobile robots. Some of these robots also have wireless communication functions that allow us to consider the concept of collaborative robots (MEFI, 2011; tinyurl.com/lf5jywj). Mobile robots can be found in a growing number of sectors, such as manufacturing, but also service industries, the health sector, and any occupation where humans accomplish dangerous tasks. Japan and Germany are the most advanced countries in the development of civilian mobile robotics, while the United States and Israel dominate the military robotics market. France's ministry of the economy estimates that the robot market could represent $30 billion by 2015 (MEFI, 2011; tinyurl.com/lf5jywj).

8. Quantum computing
This branch of computer science is still at a very embryonic stage of development but nevertheless suggests revolutionary applications in terms of calculating power and therefore security. Quantum computing uses the laws of quantum mechanics to process large volumes of information much more efficiently than traditional computing. Very specialized quantum cryptography solutions are already on the market, and some large organizations such as IBM, HP, Microsoft, Google, NASA, and Lockheed Martin, as well as startups such as D-Wave Systems in British Columbia, are investing large sums in quantum computing to accelerate the development of machines for practical applications.

9. Militarization of the Internet
In the past few years, military doctrine has changed to make control of the Internet not only an internal security issue but also a national security issue, with a sharp increase in the resources devoted to the development of offensive and defensive capabilities (Deibert, 2010; tinyurl.com/l96vzk7). At least 33 states (including Canada) have explicitly acknowledged developing offensive and defensive operational capabilities in cyberspace (Lewis and Timlin, 2011; tinyurl.com/mpfw7cv). The Pentagon spent just over $3.2 billion USD in 2012 on its defensive and offensive efforts in the cybersecurity domain (Sternstein, 2011; tinyurl.com/k9kbaes).

Six Cybersecurity Challenges

The trends outlined in the previous section will all create specific cybersecurity issues, which I examined at length in my full report (Dupont, 2012; tinyurl.com/kqqd39f). However, one important dimension to consider is their high level of integration. These nine trends are technically and socially interdependent, and some even have symbiotic relationships with each other (such as the mobile Internet and NFC payments). Other trends will converge to provide new services to individuals and businesses, such as the Internet of Things, which will benefit from scientific advances in big data to improve business productivity. Figure 1 maps a subjective sample of the interdependencies identified in the full report, and makes no claims to be exhaustive, in that new links will certainly appear as hard-to-predict disruptive innovations occur.

These interdependencies illustrate the growing number of ties linking technologies that used to be considered separately. In such a tightly coupled system, it becomes counterproductive to think about cybersecurity in narrow terms and a high-level, whole-system approach is essential in order to facilitate the emergence of effective policies and regulatory mechanisms. In this perspective, six broad security challenges can be anticipated.

1. More data
The huge quantity of information produced and stored by the vast numbers of machines that will be connected to the Internet will require the development of security technologies that remain efficient at this scale and that can detect potential risks among an ever-expanding constellation of unstructured and highly heterogeneous datasets. Given that even the smallest organizations will amass large amounts of information, the organizational capacity to keep the safe custody of such large datasets will be in question (Lane, 2011; tinyurl.com/p3y9lba).

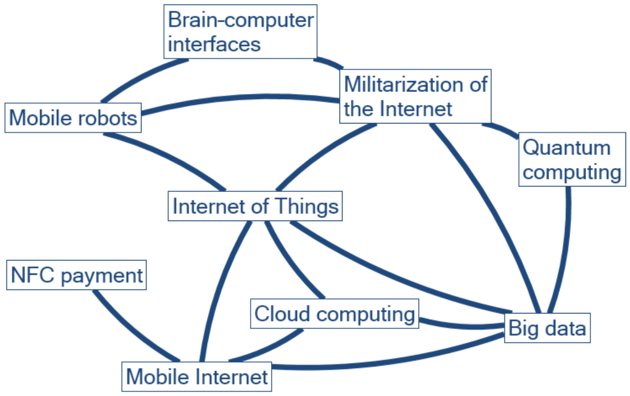

Figure 1. Key interdependencies between socio-technical cybersecurity trends

2. More connections
Each new object connected to the Internet will represent an additional entry point to the digital ecosystem that will have to be secured. This will prove particularly difficult for autonomous machines such as robots and smart meters that operate in public spaces and can be easily tampered with, or for devices that are produced in such large quantities that security features need to remain rudimentary to keep costs down (Roman et al., 2011; tinyurl.com/nqgl9qn). The proliferation of connected devices and objects will also increase surveillance capacities to an unprecedented level, and will allow malicious actors to surreptitiously collect contextual personal data that had never been available before such as geographical coordinates, on-the-fly biometric information, sounds, smells, chemical compositions, etc.

3. More movement and flows
The oceans of data generated by mobile devices, objects, and sensors will circulate in the digital ecosystem at high velocity in order to be stored, shared, and analyzed by organizations trying to discover hidden opportunities. Each movement will leave behind data traces and residues that could be exploited by malicious actors if treated carelessly. The escalation and acceleration of data flows may lead to a dilution of security responsibilities if adequate regulatory obligations are not developed and implemented.

4. More opportunities for malicious exploitation
The expansion and diversification of the digital ecosystem, which is unlikely to slow over the next decade, will benefit criminal offenders and various categories of attackers whose range of suitable targets will increase exponentially; this is a classical application of Cohen and Felson's (1979; tinyurl.com/pml7vcq) routine-activity theory. Low-skill hackers will statistically find more unprotected machines available online, while high-skill hackers will leverage these new opportunities to create larger botnets and launch more damaging and unpredictable attacks.

5. Less security by design
The "security by design" movement, which was initially inspired by C. Ray Jeffery's (1971; tinyurl.com/pdzyvox) work on crime prevention through environmental design, has now expanded beyond buildings and spaces to include objects, machines, and applications. Com-

panies and inventors are encouraged to consider how they can reduce offender opportunities early enough in the design process by undertaking research, observing users, and interviewing stakeholders (Alliance Against Crime, 2011; tinyurl.com/p3v9t2d). This long and complex method, which can be applied to software design (Gegick and Barnum, 2005; tinyurl.com/oy5fe7e), is unfortunately incompatible with faster innovation cycles where new products are brought to market as soon as possible to prevent competitors from achieving a dominant position. In contrast with other industrial sectors such as car, plane, or toy manufacturing, very limited enforceable security standards are in place to offset the absence of economic incentive in marketing safe products.

6. Less control

The growing complexity inherent to a digital ecosystem relying on highly diversified technologies that were not necessarily developed to be used by such a large proportion of the population or in an integrated manner creates technical and regulatory challenges that delay the implementation of effective control mechanisms. Legacy technical protocols and existing government institutions are not well prepared to deal with this new reality and apply industrial-era answers to digital-era problems. The case of privacy is a good example. The traditional privacy-control mechanisms that organizations, individuals, and regulatory authorities currently have available become particularly difficult to use, if not obsolete. This is because the mix of big data, cloud computing, mobile Internet, NFC payments, and the Internet of objects technologies will automatically and constantly generate huge personal data streams shared by a myriad of organizations. In such an environment, how can one ascertain what types of data are collected and retained, with what degree of accuracy and reliability, or what data retention, exchange, marketing, and destruction policies are implemented? Moreover, every disruptive technology causes the appearance of new actors in the digital ecosystem. From a cybersecurity perspective, this instability makes coordination efforts more difficult by constantly introducing new organizational actors whose abilities and willingness to contribute to the security of the ecosystem as a whole are difficult for their partners and the regulatory authorities to assess and mobilize.

Regulatory Options to Increase the Digital Ecosystem’s Resilience

While computer scientists are actively working on technical fixes to solve the six challenges listed above, the nature of the debate among social-science scholars has been much more cautious and skeptical. Efforts to design and implement regulatory mechanisms that could enhance the safety of online users have more or less explicitly been associated with governmental attacks eroding the Internet’s core values of freedom and openness (Zittrain, 2009: tinyurl.com/qb9blmc; Deibert and Rohozinski, 2010: tinyurl.com/l96vzk7; Mueller, 2010: tinyurl.com/qdgkqbx; Palfrey, 2010: tinyurl.com/m5sxsja). So, while the digital ecosystem is expanding and integrating, regulatory theory remains fragmented and reluctant to offer new alternatives to address existing and future cybersecurity challenges. If we return to the diagram of interconnected trends (Figure 1), we would ideally need to map a corresponding diagram representing links between regulatory regimes that should be reflecting these changes. This second diagram would represent the linkages that should be established between various fields of regulation (such as banking regulations, health, law, and medical ethics – for brain computer interfaces, criminal law, traffic regulations – for mobile robots, the law of war, privacy regulations, international industrial and security standards, telecom regulation, etc) in order to move toward a regulatory model that could harness this plurality instead of being constrained by it.

The concept of “regulatory pluralism” recognizes that regulation has become dispersed and that many institutions (including private actors) and tools beyond the state from a broad range of fields can be mobilized to achieve outcomes aligned with the public good (Grabosky, 1995; tinyurl.com/lk7vkkp). What characterizes regulatory pluralism is the belief that, by relying on diverse, complementary, and self-reinforcing regulatory instruments, policies can be implemented in a manner that is more responsive to the specific context, resources, and constraints of a particular sector (Crawford, 2006; tinyurl.com/lshb4wv). In other words, regulation becomes focused on hard problems to solve and outcomes to achieve instead of being obsessed by compliance to a narrow set of prescribed behaviours. In his influential book on cyberspace regulation, Lessig (2006; tinyurl.com/lf5zrfb) outlines four types of regulatory constraints that can be leveraged separately or in combination to tackle complex problems: the law, social norms, market forces, and technological architecture. Countries such as Japan, South Korea, Australia, and Germany are already experimenting with this regulatory pluralism approach to combat botnets by forging alliances of state regulators, Internet service providers, and anti-virus companies to persuade (or in some instances compel) computer users to clean their infected machines (Dupont, 2013: "An International Comparison of

Anti-Botnet Partnerships", Public Safety Canada: Ottawa). Without any need to legislate and with limited public monies, these initiatives are achieving promising outcomes through innovative regulatory approaches that harness the four levers described by Lessig.

The question then becomes how to ensure optimal and consistent participation when regulation is entrusted to a large extent to private actors. In other words, can a diversity of actors operating in a pluralistic regulatory environment be effectively incentivized and choreographed without building a large counterproductive bureaucracy (Grabosky, 2012; tinyurl.com/mwrbf8t)? In trying to answer this question, Ayres and Braithwaite (1992; tinyurl.com/muyrh78) suggest that the concept of "responsive regulation" may offer an innovative and cost-efficient alternative to the dichotomy of state-regulation versus self-regulation. Their theory rests on the core principle of the "benign big gun", where escalating enforcement practices are deployed in order to individualize the regulatory activity's intensity to the regulated actors' behaviour. The default strategy in this context is non-intrusive and delegated regulation, which is more likely to generate cooperation and innovation among private actors by allowing them discretion in deciding how best to achieve regulatory goals. For private actors that are unwilling or unable to implement effective strategies (i.e., in a case of market failure), the state retains the ability to escalate its level of interventionism by shifting to command-and-control regulations that involve various forms of punishment. Responsive regulation principles are inherently compatible with the need to preserve the innovative potential of Canadian companies in a highly competitive business context, by letting key stakeholders find optimal solutions suited to their particular needs and capacities before state interventions become escalated to more coercive and costly approaches.

Conclusion

The gap between the anticipated evolutions of the digital ecosystem and the regulatory tools that are being currently forged by regulatory authorities seem to comfort Killias' (2006; tinyurl.com/m6qmdz5) general theory of crime and security breaches. However, the major difference with Killias' historical examples of mass production of spirits, consumer goods, or the emergence of the banking system, is that the current wave of techno-social innovations is unfolding on many different fronts and that the resulting interdependencies introduce an unmatched level of complexity. We tend to think about new trends in isolation; in this article, I argue for a more holistic approach. I have sketched how nine techno-social trends will shape the digital ecosystem, and how new cybersecurity challenges and requirements will emerge as a result. Without a concerted and integrated regulatory strategy to guarantee the security and stability of the digital ecosystem, Canada's technological capacity may erode and fall behind its global competitors. Some countries such as Australia, Japan, and Germany, among others, are already experimenting with multi-stakeholder or nodal regulatory schemes to manage complex digital risks such as botnets (Dupont, 2013: "An International Comparison of Anti-Botnet Partnerships", Public Safety Canada: Ottawa). Other fields of regulation have also witnessed the emergence of ingenious initiatives that may also be transferrable to the cybersecurity environment (Braithwaite and Drahos, 2000; tinyurl.com/msp2xu5). But, expecting that the status quo or laissez-faire solutions will miraculously produce enhanced cybersecurity in this fluid environment is clearly not a sustainable option.

Acknowledgments

This article is based on two reports prepared for Public Safety Canada's National Cybersecurity Directorate. The author wishes to thank Public Safety Canada for the support provided to complete this research.

About the Author

Benoit Dupont is the Canada Research Chair in Security and Technology at the Université de Montréal, where he is Professor of Criminology and Director of the International Centre for Comparative Criminology. Professor Dupont researches the coevolution of crime and technology, focusing on offences such as identity theft, bank fraud, computer hacking, and telecommunications fraud. His political science background also leads him to examine emerging cybersecurity policies and what forms of regulation can be developed to address the new risk landscape.

Citation: Dupont, B. 2013. Cybersecurity Futures: How Can We Regulate Emergent Risks? *Technology Innovation Management Review*. July 2013: 6–11.

Keywords: cybersecurity, technological trends, forecasting, emerging risks, digital ecosystem, cybersecurity policies, regulation

Securing Canada's Information-Technology Infrastructure: Context, Principles, and Focus Areas of Cybersecurity Research

Dan Craigen, D'Arcy Walsh, and David Whyte

> *"I don't want to make the wrong mistake."*
>
> Lawrence (Yogi) Berra
> Major League Baseball player and manager

This article addresses the challenges of cybersecurity and ultimately the provision of a stable and resilient information-technology infrastructure for Canada and, more broadly, the world. We describe the context of current cybersecurity challenges by synthesizing key source material whose importance was informed by our own real-world experiences. Furthermore, we present a checklist of guiding principles to a unified response, complete with a set of action-oriented research topics that are linked to known operational limitations. The focus areas are used to drive the formulation of a unified and relevant research and experimental development program, thereby moving us towards a stable and resilient cyber-infrastructure. When cybersecurity is viewed as an inherently interdisciplinary problem of societal concern, we expect that fundamentally new research perspectives will emerge in direct response to domain-specific protection requirements for information-technology infrastructure. Purely technical responses to cybersecurity challenges will be inadequate because human factors are an inherent aspect of the problem.

This article will interest managers and entrepreneurs. Senior management teams can assess new technical developments and product releases to fortify their current security solutions, while entrepreneurs can harness new opportunities to commercialize novel technology to solve a high-impact cybersecurity problem.

Introduction

The explosive growth, complexity, adoption, and dynamism of cyberspace that have enhanced social interaction and expanded our ability to productively utilize our environment have also introduced new adversarial threats and challenges to the institutions and individuals that make up our society. Ongoing threats to our critical infrastructure have resulted in substantial loss of competitive advantage and have deleteriously impacted our way of life. Cyberbullying, cybercrime, cyberterrorism, and adversarial state-sponsored activities are all examples of malevolent attributes of cyberspace. Mitigating these malevolent attributes requires an agile, legal and ethically compliant, interdisciplinary and scientifically based research and exploratory development program in cybersecurity.

The overall cybersecurity research challenge resides within a particularly complex area, being at the intersection of behavioural sciences, formal sciences, and the natural sciences. The significant adversarial component of cyberspace has led to a view that the science of cybersecurity is a science that must support reasoning about adversaries, the core components being operations research, cybernetics, and game theory. Consistent with this perspective are "nature inspired" approaches that draw upon analogies arising from immunological and biological systems. Other areas that could usefully inform a science of cybersecurity include cryptography, formal reasoning, machine learning, and composition. Our core tenant is that the cybersecurity challenge is inherently interdisciplinary and demands coordinated attention from new perspectives for the public good.

In response to Canada's Cybersecurity Strategy (2010; tinyurl.com/md7qchf), we published a report in May 2013 for the Communications Security Establishment Canada (CSEC; cse-cst.gc.ca). Our report (Craigen et al., 2013; tinyurl.com/k6khgr6), upon which this article is based, described what is required to establish a secure, stable, and resilient information-technology infrastructure. Informed by national and international strategies, roadmaps, and problem books, we presented a research context for investigating the cybersecurity challenge. In addition, we formulated a set of guiding principles to ensure the cybersecurity research program addresses the desired improvements, outcomes, and guidance stated in Canada's Cybersecurity Strategy. Constrained by the context, and satisfying the principles, we then described the specific research focus areas. Although we were specifically responding to Canada's Cybersecurity Strategy, it is our view that the context, guidelines, and focus areas are of global consequence.

Addressing the inherently interdisciplinary challenge of cybersecurity and ultimately establishing a secure, stable, and resilient information-technology infrastructure for Canada and, potentially, the world, should also be of direct interest to managers and entrepreneurs. Being a consumer or producer of enhanced cybersecurity capability presents emerging business opportunities and demands state-of-the-art management methods to ensure a diverse ecosystem is coordinated in manner that progressively addresses operational limitations and builds wealth for the collective good.

Beyond research and experimental development, we believe the context, principles, and research focus areas presented in this article are also a useful starting point for assessing and evolving management regimes that will be required to address the challenge. We also believe the material is a useful orientation for identifying new business opportunities that will arise as new interdisciplinary perspectives related to cybersecurity are better understood.

The main body of this article is composed of three complementary sections. The first section provides a summary of related work and a description of a research context for cybersecurity in order to scope the problem domain. The second section articulates a set of guiding principles that inform the nature and kinds of specific research initiatives that should be pursued. The third section identifies particular focus areas for research and experimental development that are linked to operational limitations. Note that the core components of this article (i.e., the three complementary sections) essentially capture the current contextual state within which the nine focus areas are derived and presented. The guiding principles provide suggestions on how to progress the focus areas in a productive, action-oriented manner. Finally, the conclusion summarizes important key considerations going forward when addressing the interdisciplinary cybersecurity challenge as a whole.

Given the dynamic attributes of cyberspace, we take the perspective that the focus areas will need to be updated as circumstances warrant. Through the sharing of the focus areas we hope to generate an ongoing discussion about how to achieve the end state of a secure, stable, and resilient information-technology infrastructure.

Context of Cybersecurity Research

In this section, we provide a concise and selective literature review of the material we used to set the context for establishing an appropriate and relevant research program that addresses challenges that are: i) specific to cybersecurity or ii) shared with other domains, but of particular relevance to the cybersecurity domain. In our opinion, the referenced material provides a well-considered and useful description of the cybersecurity domain.

Recent work by Mulligan and Schneider (2001; tinyurl.com/kt3f3gq) presents the view that cybersecurity should be considered as a public good. Using public health as an example, the notion of "public cybersecurity" is articulated. This is important contextually because new policy and new institutions are implied. Exploring the shift from public health to public cybersecurity, Mulligan and Schneider also provide illustrative examples that are useful for evaluating the nature of the cybersecurity domain as enlightened from this new viewpoint.

From a scientific perspective, the material is also well founded with respect to emerging research focused on the grand challenge of establishing a "science of (cyber)security" (e.g., TRUST: truststc.org; McMorrow, 2010: tinyurl.com/35h74h6; Science of Security Workshop, 2008: sos.cs.virginia.edu; U.S. Department of Homeland Security, 2009: tinyurl.com/y98ohjr). Papers by Denning (1976; tinyurl.com/l8qxamp) and Harrison and colleagues (1976; tinyurl.com/ltnzfoe) are early examples of research that would advance a science of cybersecurity. Through discussion of classes of attacks, policies, and defenses, Schneider (2012; tinyurl.com/luj9pau) references the importance of building upon existing knowledge, particularly formal methods, fault-tolerance, and experimental computer science but Schneider also acknowledges the importance of cryptography, information theory, and game theory. Interestingly, based on safety ("no bad thing") and liveness ("some 'good thing' happens"), Schneider (2012; tinyurl.com/luj9pau) and McMorrow (2010; tinyurl.com/35h74h6) suggest new techniques to express and validate security policy requirements as part of the emerging science of cybersecurity.

With a focus on technical measures for blocking cyberattacks, a U.S. Department of Homeland Security (DHS) report (2011; tinyurl.com/65udd87) adopts the human immune system as a metaphor to motivate the need for automated collective action amongst distributed systems to defend individual computers and networks. The DHS report identifies automation, interoperability, and authentication as the building blocks that underpin a five-level focus and convergence maturity model for networked environments. The DHS also describes the attributes and desired end state of a healthy cyber ecosystem (including participants within the ecosystem).

There is also clearly a strong connection between cybersecurity research and ongoing investigations concerning security analytics and measurements (Cybenko and Landwehr, 2012: tinyurl.com/kc3nm7p; Yee, 2012: tinyurl.com/lokvcs8). As stated by George Cybenko, the founding Editor-in-Chief of *IEEE Security and Privacy* and his first successor, Carl E. Landwehr, "Accordingly, we won't find the appropriate science for understanding the evolving cybersecurity landscape in the logic of formal systems or new software engineering techniques; it's an emerging subarea of game theory that investigates dynamics in adversarial situations and the biases of competing human agents that drive those dynamics." Based upon game theory, partially observable Markov decision processes and other techniques, Carin and colleagues (2007; tinyurl.com/mkf7fyw) describe a computational approach to the quantitative cybersecurity risk assessment of intellectual property in complex systems – we believe this methodology could be augmented/generalized to also address critical infrastructure protection.

Finally, from the perspective of "Reducing Systemic Cybersecurity Risk", Sommer and Brown (2011; tinyurl.com/l2nbn5r) suggest that research responses should adopt a cross-disciplinary approach that combines "hard computer science" with the need to understand social science dimensions because "information system security are achieved only by a fusion of technology and the ways in which people and organizations actually try to deploy them". Further, Dave McMahon and Rafal Rohozinski (Bell Canada and the Secdev Group: "Dark Space Report", December 2012) state that, "Current approaches to cybersecurity are ill-suited to detecting or anticipating threats, which increasingly rely on hybrid socio-technical vectors." An example of a hybrid socio-technical vector would be phishing attacks – they have a technical component, but use sociological/psychological means to induce a user to invoke malware. McMahon and Rohozinski further suggest that, "By identifying and understanding the threat agents as threats themselves, instead of only the technology as threats, we can understand and neutralize other threats before they are created".

In this section, we have provided a context for our establishing an appropriate and relevant cybersecurity research program. Next, informed by the context, a set of guiding principles is presented for responding to the cybersecurity challenge in a productive action-oriented manner.

Principles of Cybersecurity Research

This section summarizes a set of 13 guiding principles of cybersecurity research. How was this particular set of principles determined? Firstly, the IT-security best practices (tinyurl.com/l42xht7) promulgated by our organization, the CSEC, were used as a baseline to validate these principles, as they were determined. Secondly, each principle was linked to at least one key information source first cited in the research context description. These sources are produced by recognized subject matter experts and provide more detailed explanatory material. Finally, the principles were appraised collectively as a concise but comprehensive set of principles that are anchored in a careful estimation of our own experiences, baseline best practices, the context, and ongoing engagement with cybersecurity stakeholders. The principles also provide a starting point for deliberating about the multi-dimensionality of the problem domain and its interdisciplinary nature.

Context, Principles, and Focus Areas of Cybersecurity Research

Dan Craigen, D'Arcy Walsh, and David Whyte

The following are the guiding principles we have identified:

1. Coordinate research activities to systematically progress towards achieving the attributes and desired end state of a healthy cyberecosystem (including participants within the system) (DHS, 2011; tinyurl.com/65udd87).

2. Engage social-science research labs to understand the social-science dimensions of cybersecurity, thereby augmenting "hard", computer science research (Mulligan and Schneider, 2001: tinyurl.com/kt3f3gq; Sommer and Brown, 2011: tinyurl.com/l2nbn5r).

3. Focus research on promising scientific approaches that comprehensively and rigorously underpin required security policy (Schneider, 2012: tinyurl.com/luj9pau; McMorrow, 2010: tinyurl.com/35h74h6).

4. Focus research on promising scientific approaches that comprehensively and rigorously underpin the quantitative cybersecurity risk assessment of complex systems (especially critical infrastructure) (Cybenko and Landwehr, 2012: tinyurl.com/kc3nm7p; Carin et al., 2007: tinyurl.com/mkf7fyw).

5. Focus research on promising scientific approaches to automate collective action amongst distributed systems to defend individual computers and networks (DHS, 2011; tinyurl.com/65udd87).

6. Focus on research that incorporates adversaries in models and analyses of cyberspace (McMorrow, 2010; tinyurl.com/35h74h6).

7. Engage research labs to investigate cybersecurity-related research gaps and to de-risk scientific approaches and emerging technological solutions (Schneider, 2012: tinyurl.com/luj9pau; McMorrow, 2010: tinyurl.com/35h74h6; Science of Security Workshop, 2008: sos.cs.virginia.edu; DHS, 2009: tinyurl.com/y98ohjr).

8. Leverage and influence cybersecurity-related maturity models and standards when investigating difficult problems (DHS, 2011; tinyurl.com/65udd87).

9. Build upon existing knowledge that is relevant to cybersecurity (McMorrow, 2010: tinyurl.com/35h74h6; Science of Security Workshop, 2008: sos.cs.virginia.edu; DHS, 2009: tinyurl.com/y98ohjr).

10. Leverage research that addresses the challenges of "big data" as well as domain-specific challenges (U.S. Office of Science and Technology, 2012: tinyurl.com/l2pucpt; PREDICT: predict.org).

11. Leverage research that addresses the question: "What does a data scientist do? " (IBM InfoSphere; tinyurl.com/bwupcuh)

12. Leverage existing knowledge regarding ways of working, as discussed in our full report (Craigen et al., 2013; tinyurl.com/k6khgr6).

13. Carefully address the myriad of considerations (such as those pertaining to ethics) that influence and are influenced by cybersecurity (Menlo Report, 2011; tinyurl.com/mk9b44a).

In this section, we summarized a set of 13 guiding principles of cybersecurity research. In the next section, we present the focus areas of cybersecurity research that are constrained by the context outlined in the previous section and satisfy the principles outlined above.

Focus Areas of Cybersecurity Research

The following sub-sections describe nine focus areas for cybersecurity research. To identify these focus areas, the authors assessed key research-program descriptions related to cybersecurity, which we used as a baseline to validate each focus area. Next, based upon our own expertise and experience, we ensured that each focus area corresponds to operational limitations. Finally, the focus areas were appraised by organizational stakeholders as a concise but comprehensive set of focus areas that are anchored in a careful estimation of our own experiences and ongoing engagement with cybersecurity stakeholders. Further details and a more complete list of challenges and research topics, can be found in our full report to the CSEC (Craigen et al., 2013; tinyurl.com/k6khgr6). In the sub-sections that follow, we briefly describe each of these nine focus areas as action-oriented statements accompanied with a short explanation and example challenges.

1. Improve the management and quality of signatures
A signature is a distillation of a pre-configured malicious pattern. Signatures are widely used, for example, to tersely identify cyberthreats and thereby identify and detect the activity of known malicious networks and hosts (e.g., viruses). Challenges include prioritization

and arbitration of generated events from computer network operations, false-positive reduction, and automated signature generation based on a corpus of data. Responding to the challenges will improve the detection, quality, effectiveness, complexity, fidelity, and timeliness of signature-based techniques.

2. Increase effort on anomaly detection and support discovery of new threats
Anomaly detection refers to activity that does not conform to expected behaviour or usage patterns. From a cybersecurity perspective, for example, anomalous traffic patterns in a network could suggest that a system has been penetrated and sensitive data is being exfiltrated. Challenges include specification-based intrusion techniques, data mining to support anomaly-based detection hypotheses, and mimicry-attack detection. Responding to the challenges will target new areas where anomaly detection and discovery can be explored (e.g., protocol semantics, applied mathematics, statistics, machine learning), coupled with novel techniques to minimize post-detection analysis requirements, etc., thus materially improving this field.

3. Reduce time to action through streaming and event-driven analytics
Streaming analytics refers to the inline analysis of data (e.g., Internet protocol packets, stock trades, currency trading, health monitoring) to rapidly and intelligently respond to evolving situations, potentially in near real-time. There is a spectrum of algorithms, ranging from near real-time algorithms supporting almost instant response to adversarial situations, through to longer-term algorithms that require an almost forensics-like, perspective. Identifying this algorithmic taxonomy is a research challenge in its own right. Example challenges include automated, machine-driven signature detection and near real-time correlation of events.

4. Provide dynamic defence at the network edge and beyond
A network edge is the location where the processing and enforcement of organizational policies commences. This challenging problem focuses on developing dynamic defence techniques that can rapidly interdict network attacks, using both network and host-based capabilities. The "end goal" for dynamic defence can, in fact, be two-fold: i) to mitigate the degree of damage attributed to a detected compromise by adapting the network or host environment in a timely fashion to actively resist or repel an ongoing attack, and ii) to ensure that mission-critical services are available to clients even when the network or hosts are under attack.

5. Investigate secure cloud-based systems including virtualization
Cloud computing is the delivery of computing resources over a network. Cloud computing brings challenges pertaining to scale, security, and privacy. Challenges arise from the evaluation, architecture, and design of such systems. Furthermore, there are specific concerns about contagion of malware infections across virtual instances and into the underlying base image. Virtualization is a key technology underpinning cloud computing. Accordingly, software as a service (SaaS), infrastructure as a service (IaaS), and platform as a service (PaaS) present both attractive cost savings in addition to potential security concerns (e.g., separation of virtual machines, secure application programming interfaces, authentication, secure auditing, as well as multi-latency and hypervisor vulnerabilities).

6. Investigate secure supply chains
Commercial off-the-shelf (COTS) products are those products that are commercially available, leased, licensed, or sold and do not require specific maintenance/modification. COTS products tend to vary in quality, yet also evolve quicker and more usefully in response to broader market forces. The challenges pertain to evaluation, architecture, and design, identification of security requirements, and the specification of such systems. There is a significant challenge to scale system evaluation and design to mitigate threats arising from specific products. The supply chain is of particular concern with COTS products.

7. Investigate practical enterprise-level metrics
Enterprise-level metrics allow us to answer questions that are fundamental to investment and deployment decisions, such as: "How secure is my organization?" and "How has my security posture improved through the last set of updates?" To properly manage our systems, scientifically based metrics and measures are required. Any underpinning "science of cybersecurity" will require a family of justified measures and metrics. Currently, there are no universally agreed-upon methodologies to address the fundamental questions of how to quantify system security.

8. Investigate secure mobility (including wireless)
Mobile devices are trending towards ubiquity and there is a strong desire to use capabilities available at home within the workplace, as in "bring your own device" (tinyurl.com/k5mc7th). Mobility raises unique questions from the perspective of threat risk assessment and adds potential attack vectors due to the use of wireless and

other over-the-air communication mechanisms. Challenges pertaining to evaluation, architecture, and design, identification of security requirements, and the specification of such systems once again arise, although within a different context.

9. Continuously leverage research related to the science of cybersecurity

Here, science is viewed as knowledge that results in correct predictions or reliable outcomes. Successful progress on this capability gap will provide significant science-based foundations for our cybersecurity techniques, including a deeper understanding of the interdisciplinary nature of cybersecurity. Though there are sub-areas that are solidly grounded in mathematics (e.g., formal methods and cryptography), much of cybersecurity is based on pre-scientific reasoning. Nicol and colleagues (2012: tinyurl.com/m7ufltk) have identified five hard problems relating to the science of cybersecurity: i) scalability and composability; ii) policy-governed secure collaboration; iii) security-metrics-driven evaluation, design, development and deployment; iv) resilient architectures; and v) understanding and accounting for human behaviour.

These nine focus areas have been informed by our specific experiences, but also by other international research programs. The first four focus areas concern the detection, analysis, tracking, and mitigation of cyberthreats; the subsequent four focus areas concern the means to create trustworthy systems. The last focus area effectively underpins the previous eight by arguing for a science of cybersecurity. We believe that, together, these nine focus areas provide a grounded and useful starting point for establishing a mature and unified research program that effectively addresses the overall cybersecurity challenge.

Conclusion

Here and in our full report to the CSEC (Craigen et al., 2013; tinyurl.com/k6khgr6), we have described the major components of a cybersecurity research program to secure Canada's information-technology infrastructure. Other relevant considerations that are outside the scope of this article include legal and ethical concerns, required skill sets, methods of assessing progress in science, and technology transfer within the cybersecurity domain.

Making the cybersecurity research program public offers benefits to entrepreneurs and managers of existing organizations, both large and small. Entrepreneurs can use the information to identify and act upon gap-filling and disruptive opportunities for the purpose of creating wealth. Managers of existing organizations will be able to search for ways to reduce risk and answer a myriad of questions about how to reduce costs, increase revenue, and enable their organizations to do things they cannot do today.

Moving forward, what is an appropriate path to take, given that cybersecurity must be achieved for the public good and that the challenge itself transcends any one organization? Given the key considerations just mentioned and the interdisciplinary nature of cybersecurity, we hope to establish a not-for-profit institute to bring together cybersecurity venture stakeholders and fully integrate a national research and commercialization program. The research context, principles, and focus areas described in this article will form the basis of the institute's combined research and commercialization program. And, with the help of the institute, innovative companies will be launched to provide cybersecurity solutions that address domain-specific information-technology infrastructure protection requirements that have been identified by cybersecurity stakeholders who are part of the ecosystem. The instute will function as a state-of-the-art social enterprise, ensuring that priority requirements are addressed incrementally for the public good.

In this article, we have presented a collection of cybersecurity research focus areas. Although these focus areas are well-informed by our own expertise, experiences, research, and engagement with cybersecurity stakeholders, they should be viewed as a starting point for a unified cybersecurity research and experimental development program. Given the complex aspects of cybersecurity research – due to it residing in the intersection of behavioural sciences, formal sciences, and natural sciences – it is impossible for any one organization, no matter how well informed, to fully grasp the challenges and potential opportunities. We hope that, by publishing this article and the full report, a discussion will ensue within government, academia, and industry, leading to an evolving set of cybersecurity focus areas where discoveries will result in meaningful advances towards a stable and resilient information-technology infrastructure.

Context, Principles, and Focus Areas of Cybersecurity Research

Dan Craigen, D'Arcy Walsh, and David Whyte

About the Authors

Dan Craigen is a Science Advisor at the Communications Security Establishment Canada (CSEC). Previously, he was President of ORA Canada, a company that focused on High Assurance/Formal Methods and distributed its technology to over 60 countries. His research interests include formal methods, the science of cybersecurity, and technology transfer. He was the chair of two NATO research task groups pertaining to validation, verification, and certification of embedded systems and high-assurance technologies. He received his BScH in Math and his MSc in Math from Carleton University in Ottawa, Canada.

D'Arcy Walsh is a Science Advisor at the Communications Security Establishment Canada (CSEC). His research interests include software-engineering methods and techniques that support the development and deployment of dynamic systems, including dynamic languages, dynamic configuration, context-aware systems, and autonomic and autonomous systems. He received his BAH from Queen's University in Kingston, Canada, and he received his BCS, his MCS, and his PhD in Computer Science from Carleton University in Ottawa, Canada.

David Whyte is the Technical Director for the Cyber Defence Branch at the Communications Security Establishment Canada (CSEC). He is CSEC's technical lead responsible for overseeing the implementation of the next-generation cyberthreat-detection services for the Government of Canada. He has held many positions over the last 16 years within CSEC that span both the Signals Intelligence and Information Technology Security mission lines. David holds a PhD in Computer Science from Carleton University in Ottawa, Canada. The main focus of his research is on the development of network-based behavioural analysis techniques for the detection of rapidly propagating malware.

Citation: Craigen, D., D. Walsh, and D. Whyte. 2013. Securing Canada's Information-Technology Infrastructure: Context, Principles, and Focus Areas of Cybersecurity Research. *Technology Innovation Management Review.* July 2013: 12–18.

Keywords: cybersecurity, cyberinfrastructure, cyberdefence, research, information-technology infrastructure, experimental development program, management, entrepreneurship, Canada

Peer-to-Peer Enclaves for Improving Network Defence

David W. Archer and Adam Wick

> “*If we were sincerely looking for a place of safety, for real security and success, then we would begin to turn to our communities.*”
>
> Wendell Barry
> Author, critic, and farmer

Information about cyberthreats within networks spreads slowly relative to the speed at which those threats spread. Typical "threat feeds" that are commercially available also disseminate information slowly relative to the propagation speed of attacks, and they often convey irrelevant information about imminent threats. As a result, hosts sharing a network may miss opportunities to improve their defence postures against imminent attack because needed information arrives too late or is lost in irrelevant noise. We envision timely, relevant peer-to-peer sharing of threat information – based on current technologies – as a solution to these problems and as a useful design pattern for defensive cyberwarfare. In our setting, network nodes form communities that we call enclaves, where each node defends itself while sharing information on imminent threats with peers that have similar threat exposure. In this article, we present our vision for this solution. We sketch the architecture of a typical node in such a network and how it might interact with a framework for sharing threat information; we explain why certain defensive countermeasures may work better in our setting; we discuss current tools that could be used as components in our vision; and we describe opportunities for future research and development.

Introduction

Current approaches to network defence are limited in scale and speed by the limited availability of skilled human operators and the inability of these operators to share information and work at cyberspeeds. We believe that network defence can be scaled out and enhanced through timely sharing of relevant information about common threats that are reasonably expected to affect a host in the near term. The goal is for critical information about relevant threats to be shared rapidly enough that the information is useful to a recipient in preparing a timely defence that adapts to current threat conditions. Unfortunately, in current practice, such timely and relevant sharing does not typically occur.

Current approaches to sharing threat information make use of Internet-wide threat feeds that are commercially available. Such feeds typically propagate threat information with delays on the order of minutes to hours, though Microsoft’s Cyber Threat Intelligence Program under Windows Azure promises updates as often as every 30 seconds (tinyurl.com/mslnv2h). In contrast, the attacks being reported may move from one host to another in milliseconds. In addition, commercial threat feeds report on a wide variety of threats, requiring that consumers filter and prioritize threat information before acting on it. Thus, a rapid, autonomous improvement in defensive posture against imminent threats is currently prevented both by delay in the availability of threat information and by delays due to the filtering and prioritizing of that information.

In this article, we articulate a novel design pattern for defensive cyberwarfare: *enclaves* of cooperating hosts that use autonomous, timely, peer-to-peer sharing and exploitation of relevant threat information to solve these (and other) network defence problems. We begin with an overview of our approach and its benefits, and follow with a description of current technologies that suggest our approach is viable. Finally, we call on the network-defence research and development community to improve upon and realize this vision in practice.

Proposed Approach

Enclaves can be small or large, and both intra- and inter-organizational. Individual sub-nets may form enclaves, as may corporations with similar threat profiles. Key aspects of these enclaves are that they are opt-in and peer-to-peer. Thus, nodes may dynamically change their enclave membership (and thus the threat information they receive) to get best data possible. Because enclaves are peer organizations, no central clearing-house serves as a single point of failure for an enclave. Once threat information is shared, peer hosts can use it to improve their defensive posture. In the short term, a defensive response might involve the application of simple rules. For example, if a threat against a particular piece of software is detected, instances of that software can be taken offline or more intensive defences can be deployed around it. In the longer term, defensive responses might attempt to infer the intent of adversaries or take more nuanced action. To operate in such an enclave, hosts must be able to detect threats, communicate those threats, authenticate threat data received from peers, and make use of that authenticated information. In this section, we describe our approach to sharing threat information within enclaves and how it achieves these goals.

A notional architecture of a peer agent in such an enclave is shown in Figure 1. The core of the peer agent is the Inference Engine shown at centre. This engine receives threat information from local network and host intrusion-detection systems (HIDS and NIDS, shown at left in the figure). Threat information may also be provided by additional information-gathering systems, such as the Chaff Controller, shown at bottom, which creates virtual machines on the local network to confuse attackers and gather information about their attacks. We explain more about network chaff in the section on countermeasures. The Inference Engine uses this information to control local-host countermeasures such as account restrictions or file backup, network-adapter countermeasures such as obscuration of the local-host network signature, and other mechanisms such as network chaff generation. As part of its work, the Inference Engine sanitizes locally gathered threat data and passes it to a publishing agent (part of the Pub/Sub adapter), along with contextual information that may help peers to make sense of the threat. In turn, the publisher sends this threat information to peers in the enclave. Subscribed threat data from other peers is received at the network adapter (top) and processed by the subscribing agent (also in the Pub/Sub adapter), and finally, the data is sent to the Inference Engine for interpretation and use.

Our notional peer agent is autonomous; it operates independently of human administrators and centralized server control. We propose autonomy because of the increasing disparity between the size of modern networks (and the frequency of attacks) compared to the number of trained human network analysts available for network defence (Fung, 2013; tinyurl.com/bc7nb6l). In addition, typical network-wide threats are capable of propagating faster than humans can respond (Moore et al., 2003; tinyurl.com/koweuj5). Thus, *autonomous* defensive operational elements that can be deployed in high volume, that make limited decisions, and that react at “cyberspeed”, are critical components in network defence.

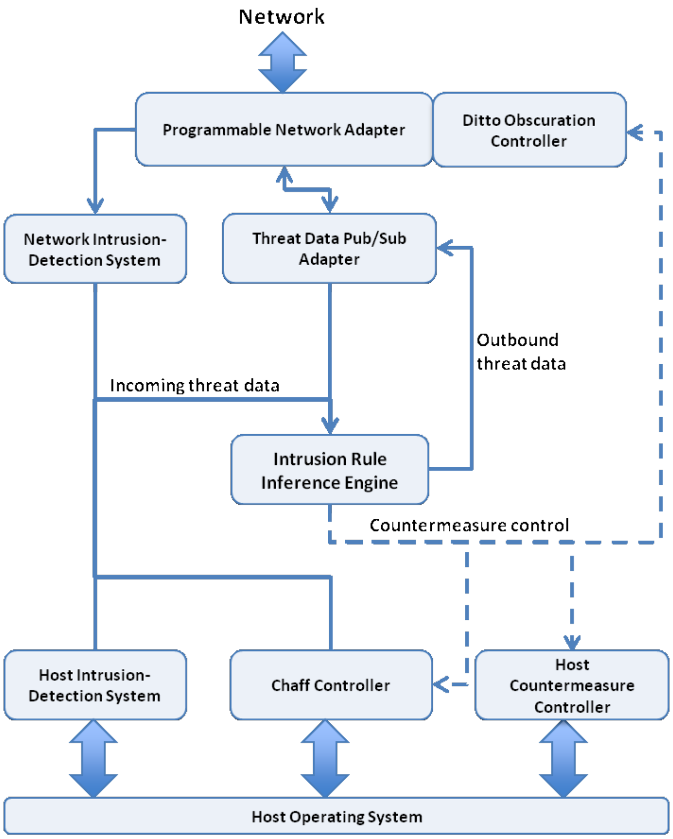
Network
Programmable Network Adapter
Ditto Obscuration Controller
Network Intrusion-Detection System
Threat Data Pub/Sub Adapter
Outbound threat data
Incoming threat data
Intrusion Rule Inference Engine
Countermeasure control
Host Intrusion-Detection System
Chaff Controller
Host Countermeasure Controller
Host Operating System

Figure 1. Notional architecture of a peer agent in a peer-to-peer defence enclave.

Peer-to-Peer Enclaves for Improving Network Defence

David W. Archer and Adam Wick

Central to our approach is the timely sharing of threat information among hosts. There is a significant prevalence of cyberattacks in which hosts sharing a network or other resource are subject to the same attack in rapid succession. This prevalence may result from a frequent structural choice in the current Internet, where subnets tend to contain hosts running similar operating systems with similar application loads (Chen et al., 2003: tinyurl.com/kw56ckh; Abu Rajab et al., 2005: tinyurl.com/l5yvzem). As adoption of IPv6 continues to expand, these problems may get worse because system administrators organize their machines into logical sub-nets that are globally addressable. A variety of advanced persistent threats (APTs) typify this iterative attack pattern. Hutchins and colleagues (2011; tinyurl.com/8qhsj5u) note that, "APT actors, by their nature, attempt intrusion after intrusion." For example, RSA Security's well-known network breach in 2011 (tinyurl.com/mvk2yjh), which started with a phishing campaign targeting two groups of employees, subsequently targeted many hosts on the RSA corporate network. Similar behaviour is seen in attacks that affect multiple networks that share characteristics of interest to an attacker. For example, Operation Aurora attacked several technology and defence corporations in 2009 (tinyurl.com/np89339), methodically exploiting the software configuration management subsystems on hosts throughout target networks. The literature shows that rapid sharing of information about such threats can be an effective enabler of improved defensive posture. For example, Weaver, Staniford, and Paxson (2004; tinyurl.com/kaxwhu3) show that defence against scanning worms can be improved by rapid communication of threat information among autonomous defensive elements.

Enclaves use a *peer-to-peer* paradigm rather than a client-server approach for sharing threat information. By peer-to-peer sharing, we mean sharing performed autonomously by participating hosts, thereby avoiding human intervention or use of a central network resource. Our motivation for this choice is that centralized resources such as server-deployed enterprise applications are attractive targets for attack (tinyurl.com/q6xyuhw), and security applications are especially attractive targets (tinyurl.com/m7zk9dn). In addition, the client-server approach requires an explicit build-out of server resources as well as investment in system administration effort, while peer-to-peer resources scale naturally as new peers are added, and they require no central system-administration resources. In addition, peer-to-peer architectures are more robust than client-server architectures due to lack of single points of failure (Lua et al., 2005; tinyurl.com/kygjjen). Conversely, peer-to-peer systems have inherent security weaknesses, because each peer is controlled by the host on which it runs. Thus, "bad actors" – peers providing irrelevant or distracting threat information to peers in an enclave – can adversely affect peer-to-peer networks more easily than client-server arrangements. We recognize that our proposal of a peer-to-peer approach requires care in authenticating and trusting peers, and we address this problem in the next section.

Our approach limits timely sharing to threat information that is likely to be immediately relevant to peers, because we expect the reasoning capability in autonomous cyberdefence elements to be limited. Our goal is that shared information should be actionable without substantial filtering, interpretation, or prioritization. For example, if a threat manifests a port scan, relevant information shared among peers might specify the ports scanned, the operating system and version of the attacked host, and the applications installed at the scanned ports on that host. A peer host might exploit this information for example by applying simple rules to block the reported ports for a specified time period if the host was running the same operating system as reported.

Component Technologies for Peer Agents

In this section, we break down our envisioned system into five concrete, manageable components: a detection system, a communication language, opt-in communication channels, secure authentication and trust mechanisms, and dynamic countermeasures. In each subsection, we describe the existing technologies that may begin to meet the needs of these components.

Detection

We expect enclaves to leverage existing host and network-intrusion-detection systems, as shown in Figure 1. Host Intrusion-Detection Systems (HIDSs) look for internal changes to a system; examples include Tripwire (tinyurl.com/d4pty), which monitors file changes, and OSSEC (ossec.net), which checks system logs and registries, and looks for rootkits. More traditional anti-virus tools, such as Norton Internet Security (tinyurl.com/23shn7p), also may be considered in the HIDS category. Network Intrusion-Detection Systems (NIDSs) such as Snort (snort.org) detect bad behaviour by sniffing packets on attached networks. Other technologies such as firewalls may also detect and report threats in a timely way.

Peer-to-Peer Enclaves for Improving Network Defence

David W. Archer and Adam Wick

Communication language

Communicating threat information among peers requires that both sender and receiver use the same language. The semantics of such a language can be captured in one or more ontologies, whereas syntax can be captured in a language specification. An ontology in this context is a machine-usable specification of the entities, concepts, and relationships in a domain of discourse. Orbst, Chase, and Markeloff (2012; tinyurl.com/kbrrhrf) describe the development of ontologies for cybersecurity at the MITRE Corporation (mitre.org) as part of an effort called Structured Threat Information eXpression (Barnum, 2013; tinyurl.com/kdov4c8). Assured Information Security (ainfosec.com) is developing an ontology for describing malware behaviour and cyberenvironments (Taylor and Hall, 2013; tinyurl.com/m5yplkz).

Communication channels

The channel for transmitting timely, relevant cyberthreat information must be: *decentralized*, to make it difficult to attack and more robust than a single point of failure; *reliable*, to ensure that threat information is delivered; *timely*, to enable peers to react at cyberspeed; and *efficient*, to minimize impact to normal business logic. Publish-subscribe middleware, such as implementations of the Data Distribution Service (DDS; portals.omg.org/dds/), are designed with such properties in mind, and thus may be suitable choices for communication among enclave members. DDS family members are fully distributed without need for brokering of mediation between publishers and subscribers. Reliability and timeliness have been demonstrated in several DDS implementations such as OpenSplice Community (tinyurl.com/p8pw24g) and OpenDDS (opendds.org).

At least one communication channel is in development specifically for transport of cyberthreat information: the TAXII sharing service (taxii.mitre.org) being developed in conjunction with MITRE’s STIX language. DDS or TAXII are existing technologies that demonstrate how the content-distribution mechanisms we envision are both feasible and practical.

Authentication

A fundamental issue in communicating threat information is the degree to which a consumer of the information should trust what is communicated. Establishing trust requires action on at least two levels: authentication of transmissions, and trust in their contents. Communication and authentication standards for data transmission are well understood in general. We expect that typical protocols such as the Secure Sockets Layer (SSL; tinyurl.com/c9jdg), or similar protocols that achieve efficient data transmission and encryption may be sufficient. Message authentication and other techniques may also be applied to authenticate threat data.

Enclave peers will need to guard against malicious or broken peers, which may correctly implement data-transmission policies but may also transmit information counter to enclave interests. This problem is the subject of ongoing research in the general case, but mechanisms based on reputation systems seem a likely solution to the problem (Resnick et al., 2000; tinyurl.com/km43orc). In a reputation system, a node keeps track of reputation data from its peers. As an example, node A may keep track of threat information provided by each of its peers. If a threat reported by one peer, B, is correlated by another peer, or system-countermeasures report stating that the threat became reality, then A may increase its “opinion” of B. If a threat never materializes and no other peer mentions it, A may decrease its opinion of B. Once generated, this reputation data can be used to quickly and easily weight threat information introduced to a node. In the long run, such reputation systems may also be used to remove peers that do not provide good, relevant data to the node and to find new peers that can provide such information.

Countermeasures

Enclaves offer a unique opportunity for dynamic adjustment in defensive posture. The timely exchange of relevant threat information allows hosts to take dynamic defensive action, and then revert to less aggressive defensive postures when threats pass. In contrast, current network defence techniques rely on static defensive postures that may impose hardships on users and system administrators. For example, countermeasures that automatically block network access (in part or in full), restrict account privileges, back up or obscure sensitive data, or temporarily disable ports can disrupt business processes and reduce utilization of computing resources if used consistently. However, if deployed for short time periods surrounding an attack, such disruption can be minimized.

Additional countermeasures may be available that are suitable for short-term, dynamic deployment, but might impose too much disruption for static deployment. Through recent research at Galois, Inc. (galois.com), we demonstrated the use of virtual-machine creation on-the-fly as a network defence technique called CyberChaff. Upon detection of an imminent threat, a CyberChaff device deploys a significant number of lightweight virtual machines onto a network, with network configurations that can be tuned to appear as particular operating systems running standard sets of services. By doing so, CyberChaff has the potential to obfuscate the network structure in order to confuse attackers. In addition, CyberChaff's virtual machines can serve as honeypots (tinyurl.com/37scmk), gathering information about patterns of cyberthreats to provide greater insight into the attackers' identities, goals, and preferred attack patterns. The Chaff Controller, shown in Figure 1, illustrates how CyberChaff fits into our notional enclave peer architecture.

Other recent research at Galois demonstrated a network stack called Ditto, which can allow a host to falsely display its configuration to external network scans. Using Ditto, a host can appear to be running a different operating system than actually used by the host. Ditto is intended to solicit attackers to waste time by applying exploits that are less likely to succeed because they target incorrect operating systems. The Ditto Obscuration Controller, shown in Figure 1, illustrates how Ditto fits into our enclave peer.

There is increasing interest in using software-defined network routing such as that provided by OpenFlow (openflow.org) for intrusion response. OpenFlow allows hosts to specify policies that classify traffic as belonging to specific network flows and thus enables redirecting of that traffic upon detection. For example, OpenFlow policies might re-direct port scanning traffic from its intended destination to a honeypot. The FRESCO framework (Shin et al., 2013; tinyurl.com/n2z24wv) is a recent system that employs a related approach. Software-defined networking might be included as part of the Programmable Network Adapter shown in Figure 1.

Conclusion

Current approaches to network defence rely on static end-point defensive postures taken by individual hosts that lack timely and relevant information about threats they may soon face; or actions orchestrated by centralized command-and-control systems that receive threat information and adjust postures slowly relative to attacks. Our vision is to change this defensive landscape by enabling the creation of enclaves that are responsive, informed, and armed. In such enclaves, each host dynamically adjusts its own defence at cyberspeed, and all hosts share information about emerging threats with their peers in a timely way. In doing so, hosts can reduce disruption to users and system administrators because some countermeasures can be deployed dynamically in response to such information instead of statically, and hosts gain the advantage of access to new countermeasures specifically designed for such dynamic deployment. Such enclaves may be localized to a single network or may include hosts from distinct networks owned by organizations that face common cyberthreats. For example, as the Internet of Things (tinyurl.com/5qr2nq) emerges and home networks grow to be more attractive targets, home networks in a physical neighbourhood may face common threats such as drive-by network hacking, and these networks may form enclaves in response.

In this article, we presented a notional architecture for hosts capable of operating in the enclaves we describe, as well as a notional means for these hosts to communicate timely, relevant threat data. For the most part, the key technologies required to create a first generation of such enclaves already exist. However, some key technologies still require advancement, and the pieces must be combined into an integrated whole. We note, in particular, the need for practical, rapid methods for describing and communicating threat information, as well as the need to develop advanced-decisions engines capable of receiving, analyzing, and acting on network threats.

Peer-to-Peer Enclaves for Improving Network Defence

David W. Archer and Adam Wick

About the Authors

David Archer is a Research Program Lead at Galois, Inc., where he directs research into high-assurance methods for large-scale cyberconflict. He holds a PhD in Computer Science from Portland State University in the United States as well as an MS in Electrical Engineering from the University of Illinois at Urbana-Champaign. Dr. Archer's research interests also include efficient methods for computing on encrypted data, and information integration, assurance, and provenance. At Intel Corporation, Dr. Archer was instrumental in the development of the communication network for the ASCI Red TeraFLOPS system at Sandia, and in the development of multiple generations of high-performance server and workstation memory and I/O systems.

Adam Wick directs the Systems and Networking Group at Galois, Inc., where he has worked with DARPA to create advanced network-defence techniques, including CyberChaff and Ditto. He holds a PhD in Computer Science from the University of Utah in the United States, as well as a BS in Computer Science from Indiana University Bloomington. Dr. Wick also has been collaborating with SRI, LG, and others to build secure mobile devices for the United States Marine Corps. Prior to this work, he developed the HaLVM, a lightweight machine for running custom, single-purpose applications in the cloud. In all of this work, he maintains a focus on using next-generation operating system and networking technology to create practical tools for critical systems.

Citation: Archer, D.W. and A. Wick. 2013. Peer-to-Peer Enclaves for Improving Network Defence. *Technology Innovation Management Review.* July 2013: 19–24.

Keywords: network defence, cybersecurity, enclave computing, dynamic cyberdefence, cyber countermeasures, peer-to-peer

Keystone Business Models for Network Security Processors

Arthur Low and Steven Muegge

> *"Your ability to negotiate, communicate, influence, and persuade others to do things is absolutely indispensable to everything you accomplish in life. The most effective men and women in every area are those who can quite competently organize the cooperation and assistance of other people toward the accomplishment of important goals and objectives."*
>
> Brian Tracy
> Entrepreneur, business coach, author, and speaker

Network security processors are critical components of high-performance systems built for cybersecurity. Development of a network security processor requires multi-domain experience in semiconductors and complex software security applications, and multiple iterations of both software and hardware implementations. Limited by the business models in use today, such an arduous task can be undertaken only by large incumbent companies and government organizations. Neither the "fabless semiconductor" models nor the silicon intellectual-property licensing ("IP-licensing") models allow small technology companies to successfully compete. This article describes an alternative approach that produces an ongoing stream of novel network security processors for niche markets through continuous innovation by both large and small companies. This approach, referred to here as the "business ecosystem model for network security processors", includes a flexible and reconfigurable technology platform, a "keystone" business model for the company that maintains the platform architecture, and an extended ecosystem of companies that both contribute and share in the value created by innovation. New opportunities for business model innovation by participating companies are made possible by the ecosystem model. This ecosystem model builds on: i) the lessons learned from the experience of the first author as a senior integrated circuit architect for providers of public-key cryptography solutions and as the owner of a semiconductor startup, and ii) the latest scholarly research on technology entrepreneurship, business models, platforms, and business ecosystems. This article will be of interest to all technology entrepreneurs, but it will be of particular interest to owners of small companies that provide security solutions and to specialized security professionals seeking to launch their own companies.

Introduction

New business models are needed for small suppliers of network security processors and specialized security products. The conventional business models in use today favour large, established incumbents who develop products for large and well-understood markets. Ideally, new business models would enable and reward continuous innovation by both large and small companies to produce a continuous stream of novel security products for niche markets. The beneficiaries would include the buyers of specialized cybersecurity products and their users, the technology entrepreneurs who develop and commercialize specialized security products, and the engineers and product designers with a broader range of employment and contracting opportunities.

Keystone Business Models for Network Security Processors

Arthur Low and Steven Muegge

Network security processors are specialized components of high-performance security systems used by organizations such as banks, government embassies, and multinational corporations. They provide *acceleration* of the cryptography functions that encrypt and decrypt outgoing and incoming information and protect against intrusion by adversaries. Security systems with hardware acceleration have greater performance than systems that implement the cryptography functions in software, but are more costly and require more time and specialized expertise to develop, implement, and deploy.

There are two broad categories of business models in use today for providers of network security processors and the security products that employ them. Both categories favour large multinational incumbents such as IBM (ibm.com), Hewlett-Packard (hp.com), Bull SAS (bull.com), SafeNet (safenet-inc.com), and Thales Group (thalesgroup.com) rather than small companies and new entrants. “Fabless semiconductor” models require commitment of large up-front capital, exposing investors to significant risk. Silicon “IP-licensing” models prevent the small company from interacting directly with customers and end-users, and because the customer relationship is owned by the systems integrator who packages the complete solution, small suppliers cannot easily appropriate a significant portion of the value that their innovations create for customers.

This article contributes an alternative approach that we refer to here as the "business ecosystem model for network security processors". It builds on lessons learned from the industry experience of the first author and implements concepts from the latest scholarly research on technology entrepreneurship (Bailetti, 2012: timreview.ca/article/520; Bailetti et al., 2012: 557), business models (Muegge, 2012: 545; Bailetti, 2009: 226), platforms and keystones (Bailetti, 2010; 355), and business ecosystems (Muegge, 2013: 655; Muegge, 2011: 495; Bailetti, 2010: 325; Carbone, 2009: 227; Hurley, 2009: 276; Bailetti, 2008: 138). This approach has several parts, including a *network security processor platform* that companies can use and reconfigure to build innovative security solutions for niche markets, a *keystone business model* for the company that leads platform maintenance and evolution, and a *business ecosystem* of companies that develop complementary products, services, and technologies, contribute assets to the platform, and build security products that utilize the platform. The ecosystem approach enables new business models for participating companies. Building solutions on top of the proposed platform does not require the sale of large volumes to generate profits. Moreover, it allows small companies to interact directly with end-customers and retain the rights over the intellectual property they create.

The body of this article is structured in four sections. The first section reviews the conventional business models used by providers of network security processors and discusses their weaknesses and limitations. The second section presents lessons learned from the industrial experience of the first author as a cryptography chip designer and entrepreneur. The third section builds on the lessons learned to develop the business ecosystem model for network security processors; it explains the business model of the ecosystem keystone, the technology that supports the ecosystem, and the new opportunities for business model innovation by companies participating in the ecosystem. The fourth section concludes with a renewed call for innovation in the cybersecurity domain – not only of novel technology but also of *novel business models* that fully exploit the opportunities enabled by technological innovation.

Conventional Business Models

A business model provides a concise explanation of how a business operates. Many business model frameworks have been proposed. This article employs the technology entrepreneurship framework previously published in the *TIM Review* (Muegge, 2012; timreview.ca/article/545) and employed with technology entrepreneurs in the Lead to Win ecosystem (Bailetti and Bot, 2013; timreview.ca/article/658). Although each company's business model may comprise a unique combination of customer pain points, stakeholder value propositions, a profit formula of revenues and costs, and the company's capabilities, it is often useful to identify and label groups of business models that share some similar features. The three groups of interest in this section are: i) integrated device manufacturers, ii) fabless semiconductor companies, and iii) silicon IP-licensing companies.

Prior to the 1980s, most companies that developed integrated circuit devices were *integrated device manufacturers.* Vertically integrated firms would own and control their own production facilities, including a foundry for fabricating semiconductor wafers, and perform basic research, product design, manufacturing, sales, and support – all in-house.

Keystone Business Models for Network Security Processors

Arthur Low and Steven Muegge

Fabless semiconductor business models became possible in 1987 when the Taiwan Semiconductor Manufacturing Company (tsmc.com) first offered the use of an integrated circuits fabrication facility to companies who could design their own integrated circuits. Instead of investing billions of dollars up-front to acquire and operate an integrated circuits fabrication facility, a fabless semiconductor company could acquire electronic design automation (EDA) software, employ engineers to design integrated circuits using the EDA software, and outsource the manufacturing to others. The non-recoverable engineering costs to produce a new integrated circuit must be recouped from product sales. To be profitable, a fabless semiconductor company requires high sales volumes – typically in the tens of thousands or hundreds of thousands of units.

Silicon *IP-licensing* business models require a company to license modular design units to become parts of integrated circuits designed by others. An IP-licensing company generates revenue from some combination of fixed fees per unit of intellectual property and royalties paid per device manufactured. ARM Holdings (arm.com) was the first company to successfully employ a business model with IP-licensing. ARM developed a “soft” reduced instruction set (RISC) microprocessor design that customers could license and embed within their integrated circuit designs to control applications-specific logic. The consumer electronics market grew rapidly when highly integrated microchips with embedded ARM processors enabled significant cost and size reductions. Smart, hand-held communications-enabled devices, such as cell phones, moved from science fiction to fact almost overnight. By 2012, ARM was employing more than 2000 people and ARM's partners had shipped more than 30 billion ARM-based integrated circuits (ARM Annual Report, 2012; tinyurl.com/kvgzuf6).

Despite these large-company successes, neither the fabless semiconductor models nor the IP-licensing models are appealing for small providers of security solutions – for reasons developed in the next section.

Background and Lessons Learned

The business model insights and platform architecture that enable the business ecosystem model for network security processors have evolved over the past 13 years. In 2000, Chrysalis-ITS extended its business of developing specialized hardware and software for the public-key infrastructure (PKI) market by opening a fabless semiconductor division to develop a high-performance line of network security processors as "systems on chips". Chrysalis-ITS's first system on a chip, the Luna 340, integrated five microprocessors with instruction sets extended to implement a number of important security operations, such as Internet Protocol Security (IPSec) and the RSA public-key cryptographic (PKC) algorithm. Both are used in banking networks and Internet security based on the secure socket layer (SSL) protocol. In 2001, Chrysalis-ITS introduced the Luna 510, a product that delivered 100 times greater performance than the Luna 340. One microprocessor provided SSL-protocol control and data-flow management between multiple instances of highly optimized encryption and hashing algorithm processors. In 2004, Chrysalis-ITS was acquired by Rainbow Technologies, which then merged with SafeNet (safenet-inc.com). In 2007, Elliptic Technologies (elliptictech.com) developed a public-key cryptographic algorithm compute engine. The engine was based on an arithmetic logic unit designed to flexibly compute over any integer size up to thousands of bits the modular arithmetic functions that are the basis for security applications based on public-key cryptography. In 2009, Crack Semiconductor (cracksemi.com), a company founded by the first author of this article, developed a scalable, modular architecture for optimally computing these modular arithmetic functions in a very low-cost field-programmable gate array (FPGA). The architecture was refined over several generations so that current implementations rival the performance of the Luna 510 when coupled to an embedded applications processor. The proposed platform of the business ecosystem model for network security processors is an implementation of the next generation in the evolution of this architecture.

The first author’s industry experience as a designer and entrepreneur suggests five lessons for small suppliers of security solutions, each of which is expanded upon in the subsections that follow:

1. Control the key technology components that differentiate your business from others.

2. Avoid fabless semiconductor models for small markets.

3. Go after niche markets that are unattractive to large incumbents.

4. Implement the best-available design methodologies, tools, algorithms, and architectures.

5. Look to emerging industry standards for global opportunities to innovate.

1. Control the key technology components that differentiate your business from others
Silicon IP-licensing models place the IP supplier in a subordinate role in the value chain; from a subordinate role, it is difficult to charge license fees that are high enough to recoup R&D costs. ARM has been very successful with IP-licensing for high-volume consumer devices, but comparable mass-market sales volumes are not feasible for security applications. Furthermore, the downstream systems integrator controls the relationship with customers and end-users. For these reasons, silicon IP-licensing models are not appealing for small providers of security solutions.

The Luna 510 provides a cautionary tale regarding IP licensing and loss of control of key technology components. The Luna 510 was a technological breakthrough in network security processor design, but failed to reach the market when the Luna 340 failed. Despite the viability of the Luna 510 design, investors shut down the entire semiconductor division when it became clear that the sunk costs of the Luna 340 project would produce no revenue. Furthermore, the entirely independent and original in-house development of the Luna 510 was tainted by a clause in the Luna 340 development contract with a third-party that assigned a small but meaningful right to “derivative works” to the third-party. Because the IP was “tainted” with unquantified legal issues, new investors were unwilling to recapitalize the semiconductor division as a separate company. Thus, due to factors outside the control of the development team – in particular, the failure of another product and the loss of control over intellectual property – the Luna 510 was never produced.

2. Avoid fabless semiconductor models for small markets
Fabless semiconductor models incur high R&D costs and non-recoverable engineering costs to produce a custom integrated circuit. To recoup these costs, revenues must be in the hundreds of millions of dollars, which requires in-depth market knowledge, large sales volumes of tens or hundreds of thousands of units or very high selling prices and profit margins, and venture-capital or other institutional backing. Opportunities with these characteristics are rare for small providers of security technologies.

PMC-Sierra (pmcs.com) is an example of a successful fabless semiconductor company. PMC-Sierra achieves sales in the hundreds of millions of dollars per year by providing high-performance optical-networking integrated circuits to large telecommunications equipment manufacturers such as Cisco Systems (cisco.com) and Huawei (huawei.com). Development of a new integrated circuit may cost PMC-Sierra $30 million to design, and it may incur $3 million in non-recoverable engineering charges. The integrated circuit design will be developed to a specification that meets the needs of several key clients, and features are included based on significant volume commitments. Like other companies employing fabless semiconductor models, PMC-Sierra assumes significant risk and revenue loss if the integrated circuit design is late or fails to function as specified.

3. Go after niche markets that are unattractive to large incumbents
Large incumbents employing either silicon IP-licensing models (such as ARM in the consumer products market) or fabless semiconductor models (such as PMC-Sierra in the telecommunications equipment market) *cannot* be profitable in small niche markets where their high cost structures and requirements for large sales volumes become a liability. Markets that are unattractive to large incumbents such as ARM and PMC-Sierra are an opportunity for small security providers – if those companies can be profitable at small-to-medium sales volumes.

Going after niche markets of *a thousand units* or *a hundred units* is not possible with the same technology and business models used today by incumbents; innovation is required in both the technology and business models used by small security providers.

4. Implement the best-available design methodologies, tools, algorithms, and architectures
Technology failure guarantees business model failure. Getting the technology right is necessary but not sufficient for success.

The Luna 340 network security processor is an example of what can go wrong when companies do not implement the most appropriate design methodologies, tools, and algorithms, and architectures. A team of engineers worked for several years to design and implement the Luna 340. Several early management decisions, intended to reduce costs and eliminate steps, became serious problems late in the development process. To save money on expensive EDA software licenses, a critical integrated circuit layout tool was not upgraded. Fatal circuit-timing errors were introduced, which the tool upgrade would have detected and fixed. A second design iteration – an expensive and time-consuming redesign of the integrated circuit – also failed to get the timing right. These problems were further compounded by other performance-degrading design flaws and by an inefficient architecture. In contrast, the technically successful Luna 510 employed more efficient architectures for a faster and more physically controllable hardware implementation, state-of-the-art synthesis algorithms for placement and routing, and a prototyping methodology including verification in full-speed FPGA-based prototypes, then in fabrication "shuttles" (where many companies would share a silicon wafer) before commitment to the full fabrication and manufacturing process. These design methodologies, tools, algorithms, and architectures could have been employed from the beginning of the Luna 340 project; cuttings costs in these areas was very costly later. Past research on product development has consistently found that greater early investment in architecture and flexibility results in better-performing projects (e.g., MacCormack et al., 2001; tinyurl.com/am6axfs) and the experience of the Luna 340 developers supports these findings. Greater upfront exploration of architecture and algorithms and upfront adoption of appropriate tools and prototyping methodologies could have avoided the costly delays that happened later.

For a conventional integrated circuit design, these upfront items appear as "sunk costs" to be minimized by management. However, when innovation occurs within and on top of a platform – the ecosystem approach recommended here – design methodologies, tools, algorithms, and architectures are investments in the future, to be recouped over many niche custom designs and derivative products.

5. Look to emerging industry standards for global opportunities to innovate

Small companies need to address opportunities that are global rather than local or regional (Tanev, 2012; timreview.ca/article/532), and emerging industry standards can provide insights into global opportunities. An example is the new ISA100.11a standard (isa.org/ISA100-11a) for wireless sensor networks. ISA100.11a differs from WirelessHART, a competing standard from the HART Communications Foundation (hartcomm.org), by including the *option* to use public-key cryptography technology for the provisioning of new devices joining the network. Because ISA100.11a is a new standard, and public-key cryptography is optional rather than required, few vendors are implementing this option in their first-generation ISA100.11a- and WirelessHART-compliant products. However, activity within the standards groups suggests that public-key cryptography will become increasingly important in the future: the International Society for Automation (isa.org), steward of the ISA standards, is also pursuing standardization of public-key cryptography technology in many areas, for example, to enable over-the-air (OTA) provisioning of devices. Participation in standards development can provide small security providers with valuable insights into possible futures, as well as opportunities to gain early access to information, build relationships with potential collaborators, shape requirements, and influence the technical direction of standards.

Participation in industry standards development has traditionally been a gamble for small companies using conventional business models. Costs include money and time, and the outcome is always uncertain: standards can fail for technical or political reasons, or adopters may converge on a different competing standard. However, the payoffs can be large. For example, Crack Semiconductor has developed security technologies ahead of an expected global market for wireless sensor networks for industry control (Low, 2013; timreview.ca/article/682). Furthermore, a business ecosystem approach to developing security products can substantially reduce the costs and risk of participating in standards development while retaining all the potential benefits. Participation in the development of the ISA100.11a standard is an important aspect of Crack Semiconductor's network security processor platform strategy. Other companies in the ecosystem benefit from the information and influence while sharing the costs and obligations.

In summary, for small companies of security solutions to compete successfully with established incumbents, a new approach is needed. That new approach should address global opportunities in niche markets, using the best-available design methodologies, tools, algorithm, and architectures, with business models unlike those commonly in use today by large incumbents. The next session describes one such approach.

An Alternative Approach: The Business Ecosystem Model

Business ecosystems provide a way for small companies to achieve more, learn faster, and reach farther than otherwise possible, while sharing risks and costs with others (Muegge, 2013; timreview.ca/article/655). Hurley (2009; timreview.ca/article/276) identifies several benefits enjoyed by participating entrepreneurs, including reduced barriers to market entry, increased access to customers, reduced operating costs, and the means to overcome regional limitations. Carbone (2009; timreview.ca/article/227) argues that business ecosystems can also enable business model innovation, especially by companies providing complementary assets.

Business ecosystem approaches have been previously developed for various domains, including job creation through technology entrepreneurship (e.g., the Lead to Win ecosystem: leadtowin.ca; Bailetti and Bot, 2013: timreview.ca/article/658), community development of open source software tools and frameworks (e.g., the Eclipse ecosystem: eclipse.org; Muegge, 2011; timreview.ca/article/495), and communication-enabled applications (e.g., the Coral CEA ecosystem; coralcea.ca; Pyke, 2010; timreview.ca/article/347). This article is the first known application of the business ecosystem approach to the domain of network security processors. However, the basic premise is similar to that of these other domains: ecosystem participants innovate together to solve bigger network-security problems that any one small or medium-sized company could address on its own.

As in other domains, the business ecosystem model for network security processors has several codependent parts. The most essential components in this domain are: i) a *keystone company* that owns, operates, and evolves the platform; ii) a *platform* of modular technology building blocks that others can utilize, build on, and contribute to; and iii) a *network of participating companies* that can innovate in new ways. Below, each component is briefly described in its own subsection.

Keystone business model

The keystone is the company that owns, operates, and evolves the platform (Bailetti, 2010; timreview.ca/article/355). The keystone plays a central role; for this ecosystem model to succeed, there must be a keystone business model that earns attractive profits for the keystone company.

Table 1 compares the proposed business model of the ecosystem keystone with the conventional fabless semiconductor business models and IP-licensing business models described in previous sections. The rows in Table 1 are a subset of the components of the technology entrepreneurship business model framework, selected to emphasize the salient differences. There are many similarities not shown in the table; for example, all three models are different ways of addressing the same basic "pain points" of cybersecurity.

Consistent with lesson 1, the keystone controls the key components of the technology platform – especially the cryptography algorithms, hardware acceleration, and platform architecture (described in the second subsection) – while enabling complementary innovation by other companies. Incentives are aligned, because success of the keystone business model *critically depends on* success by participating companies (described in the third subsection). Also consistent with lesson 1, participating companies keep control of their own differentiating innovations, with the option to selectively contribute specific innovations back to the platform for use by others.

Technology that supports the keystone business model

The platform that anchors a business ecosystem can take many different forms – including a product, process, location, service, or technology (Bailetti, 2010; timreview.ca/article/355). The platform for network security processors is the continued evolution of the architecture previously described in the section on background and lessons learned. It provides the essential technology components of a network security processor, tested and verified together as a system, in a modular form that can be configured in different ways, and extended with new application-specific functionality implemented in software. Cryptography functions are implemented in flexible programmable logic, avoiding the non-recoverable fixed costs of new custom silicon integrated circuits, while providing real-time performance far exceeding a software-only system on an embedded microprocessor. Thus, a new design built on the platform can be profitable at much lower sales volumes than previously possible.

The platform is made possible by an innovative network security processor architecture developed by Crack Semiconductor (cracksemi.com). The current implementation is built on the Xilinx (xilinx.com) Zynq Extensible Processing platform (EPP; tinyurl.com/kecww6k), a flexible "system on a chip" that combines a large array of programmable logic with general purpose microprocessors – more specifically, a hardened dual-core ARM-9 processor. The first microprocessor runs an SSL software library that interfaces to public-key cryptography algorithms implemented on the chip in programmable logic. The second microprocessor runs the software that provides custom requirements for specialized niche applications. The platform includes a default software stack for the second processor that includes a Linux-based operating system, a suite of open source

Keystone Business Models for Network Security Processors

Arthur Low and Steven Muegge

Table 1. Comparison of network security processor business models

	Fabless Semiconductor Business Models	**IP-licensing Business Models**	**Keystone Business Model for a Network Security Processor Business Ecosystem**
Customer	Systems integrators and providers of high-performance security products that require hardware acceleration by network security processors.	Systems integrators that provide integrated circuit products that require on-chip network security processors.	Various levels of systems integrators and demanding end-customers of high-performance security solutions. Systems integrators can participate in the ecosystem to become partners rather than customers or competitors.
Profit formula	Revenues are *product sales* of silicon integrated circuit devices. Costs include the EDA tools, R&D, and non-recoverable engineering (NRE) of outsourced integrated circuit manufacturing. For revenues to exceed costs, high sales volumes are needed (e.g., PMC-Sierra providing products to telecom equipment manufacturers).	Revenues are some combination of fixed *license fees* per unit of IP and *royalties* paid per device manufactured. Costs include the EDA tools and R&D to develop modular blocks of IP to license. For revenues to exceed costs, high sales volumes are needed (e.g., ARM providing IP for mass-market consumer electronics industry).	Revenues are *product sales* from a continuous stream of novel security products for niche markets. Products could include modular components of cybersecurity systems or complete security solutions. Costs include maintenance and extension of the platform, orchestration of innovation within the ecosystem, and investment in ecosystem health and growth. Revenues and costs are shared with participating companies.
Capabilities required	• Multi-domain experience in semiconductors and complex software security applications • Multiple iterations of hardware and software configurations	• Multi-domain experience in semiconductors and complex software security applications • Multiple iterations of hardware and software configurations	• *Platform* to be reconfigured and built on by others • *Network of participating companies* of at least three types: i) providers of security products, ii) providers of platform complements, and iii) users of security products
Applicability and context	• Favours large incumbents	• Favours large incumbents	• Attractive to small and large companies and new entrants • Enables *opportunities for business model innovation* by participating companies • Can be profitable with sales of thousands or hundreds of units

middleware, and assorted security applications. A niche application could require a custom stack that removes unneeded components, swaps out some components for specialized substitutes, and adds new proprietary custom code.

Opportunities for ecosystem companies
Participation in the network security processor ecosystem is appealing for at least three categories of company: i) providers of specialized niche technologies that complement the platform assets; ii) system integrators that build specialized security products on top of platform assets; and iii) demanding users of security products that participate in order to influence the evolution of the platform and of products that build on the platform. Examples of platform complements include hardware and software interfaces and drivers, specialized software at the middleware and application layers of the stack, and new cryptographic algorithms; providers may choose to selectively contribute some technologies and assets into the platform for use by others, for example to stimulate demand for the provider's proprietary products and services. Examples of demanding users of security products include banks and other financial institutions, governments (especially military applications and government foreign offices), institutions in the medical industry, the operators of critical infrastructure such as nuclear power facilities, and corporations. Such participants could be motivated to shape requirements, send strong signals of support, influence technical work with their investment, and gain early access to information. These motivations are similar to those for companies to participate in standards groups (lesson 5).

The network security processor ecosystem enables new opportunities for business model innovation by participating firms of all three categories identified previously (providers of complements, providers of security products, and demanding users). Returning to the components of the technology entrepreneurship business model framework (Muegge, 2012; timreview.ca/article/545), participants can: i) gain access to new capabilities; ii) reduce cost structures; iii) enable new revenue streams; iv) reach new stakeholders with new and stronger value propositions; and v) address new problem spaces that would otherwise be unavailable.

Security products developed with this approach could be profitable at sales volumes of thousands or hundreds of units – orders of magnitude below the minimum volumes required for security products using the conventional business models in use today. Providers can develop highly specialized niche products that would not otherwise be viable, for customers willing to pay high selling prices for dedicated solutions to their specialized security problems.

The network security processor ecosystem would be membership-based with restrictions and approvals required for entry. Closed membership is an important and necessary point of difference from, for example, the open ecosystems anchored around community-developed open source software where anyone can participate (e.g., Muegge, 2011; timreview.ca/article/495). The most important factor requiring this difference is government policy and regulation of cybersecurity technology: some nations regulate strong cryptography and the exchange of cryptography technology with other nations as a security concern. The United States, for example, has a body of rules including the International Traffic in Arms Regulations (ITAR; tinyurl.com/8l9zvhh), the United States Munitions List (USML; tinyurl.com/k8tvoj5), and the Arms Export Control Act (AECA; tinyurl.com/8yhb7wx), that have implications for international collaboration on cybersecurity. Some engagements may require approval from one or multiple jurisdictions. The keystone company plays a central role in developing and maintaining the membership criteria and rules of conduct, in accordance with the laws of its jurisdiction.

Conclusion

This article has argued that small innovative suppliers of network security processors and high-performance security applications that require network security processors for hardware acceleration should consider forming a business ecosystem. The configuration described here includes a platform of reconfigurable and extensible network security processor technology, a business model for the keystone company that maintains and evolves the platform architecture, and a network of participating companies that innovate within and on top of the platform. The ecosystem enables new opportunities for business model innovation by participating companies. Incentives are aligned: success of the keystone critically depends on the participation and business success of the companies that build on and contribute to the platform, including providers of niche security technologies, providers of security products that utilize the platform, and demanding end-users of security products. The outcome is a continuous stream of security innovation and of specialized security products – including products with projected sales volumes in the thousands or hundreds of units that are not economically viable with conventional business models. We call upon managers of companies large and small, and upon technology entrepreneurs seeking new opportunities, to join us in making this happen.

This ecosystem model requires some aspects of the overall solution to be shared with collaborators and partners. The platform provides a high entry barrier that protects the ecosystem from competitors, because there is no disclosure of the proprietary acceleration technology that integrates high-performance cryptographic compute offload processors with a low-level cryptographic library. Partners can therefore more rapidly develop advanced software solutions because they do not need to solve the optimization problems they would encounter if they had to develop their own network security processor. The platform's value increases significantly due to the strong network effects that are associated with multiple third-parties developing software that complements the platform.

We conclude with a renewed call for innovation in the cybersecurity domain. The technological challenges of cybersecurity have received much attention in this issue of the *TIM Review* as well as within this article. But equally daunting are the business model challenges. Just as business model innovation is required to fully exploit the network security processor platform described here, we expect that the commercial value of future innovation in cybersecurity technology may remain latent and unrealized until it is unlocked by corresponding innovation in business models and commercialization.

About the Authors

Arthur Low is the founder and Chief Executive Officer of Crack Semiconductor, a supplier of high-performance cryptographic silicon IP used in some of the most demanding security applications. Arthur has a number of patents in the field of hardware cryptography. He has worked for a number of IC startups as a Senior IC designer and Architect and gained much of his fundamental IC design experience with Bell-Northern Research in the early 1990s and with IBM Microelectronics in the late 1990s. Arthur has a BSc degree in Electrical Engineering from the University of Alberta in Edmonton, Canada, and is completing his MSc degree in Technology Innovation Management in the Department of Systems and Computer Engineering at Carleton University in Ottawa, Canada.

Steven Muegge is an Assistant Professor at the Sprott School of Business at Carleton University in Ottawa, Canada, where he teaches within the Technology Innovation Management (TIM) program. His research interests include open and distributed innovation, technology entrepreneurship, product development, and commercialization of technological innovation.

Citation: Low, A. and S. Muegge. 2013. Keystone Business Models for Network Security Processors. *Technology Innovation Management Review.* July 2013: 25–33.

Keywords: cybersecurity, business model innovation, technology entrepreneurship, commercialization, business ecosystems, platforms, semiconductors

Managing Cybersecurity Research and Experimental Development: The REVO Approach

Dan Craigen, Drew Vandeth, and D’Arcy Walsh

> *“No one means all he says, and yet very few say all they mean, for words are slippery and thought is viscous.”*
>
> Henry Adams (1838–1918)
> Journalist, historian, academic, and novelist

We present a systematic approach for managing a research and experimental development cybersecurity program that must be responsive to continuously evolving cybersecurity, and other, operational concerns. The approach will be of interest to research-program managers, academe, corporate leads, government leads, chief information officers, chief technology officers, and social and technology policy analysts. The approach is compatible with international standards and procedures published by the Organisation for Economic Co-operation and Development (OECD) and the Treasury Board of Canada Secretariat (TBS). The key benefits of the approach are the following: i) the breadth of the overall (cybersecurity) space is described; ii) depth statements about specific (cybersecurity) challenges are articulated and mapped to the breadth of the problem; iii) specific (cybersecurity) initiatives that have been resourced through funding or personnel are tracked and linked to specific challenges; and iv) progress is assessed through key performance indicators.

Although we present examples from cybersecurity, the method may be transferred to other domains. We have found the approach to be rigorous yet adaptive to change; it challenges an organization to be explicit about the nature of its research and experimental development in a manner that fosters alignment with evolving business priorities, knowledge transfer, and partner engagement.

Introduction

In many academic, private, or public contexts, research programs must address critical challenges and produce innovative discoveries. In addition, these discoveries often must be efficiently and effectively transformed into technological capabilities. Research programs that are continuously adaptive to business, technical, legal, and other drivers or constraints can enable the vitality and relevancy of research and experimental development (R&ED). Adaptive research programs can play a critical role in ensuring that major or minor scientific or technological breakthroughs respond to evolving operational environments.

In this article, we present the Research in Evolution (REVO) approach for managing a research program that we employ to address cybersecurity-related concerns. At its core, REVO is based upon distinguishing what R&ED needs to be done from what R&ED is being done. The method is intentionally compatible with the standards from the Organisation for Economic Co-operation and Development (OECD; oecd.org) and the Treasury Board of Canada Secretariat (TBS; tbs-sct.gc.ca) that provide guidance about the scope of such programs, related definitions, and performance indicators. The method is rigorous enough to enable (on-demand) reporting on scientific expenditures and personnel with respect to research, experimental de-

Managing Cybersecurity Research and Experimental Development: REVO Approach

Dan Craigen, Drew Vandeth, and D'Arcy Walsh

velopment, and related scientific activities as defined by the OECD.

REVO is not just a vehicle for producing reports. The method challenges researchers, related scientific-activity analysts, and research-program managers to be sufficiently explicit about the problem space and the solution space to enable the continuous re-alignment of scientific or technological investigations based upon a collective understanding of what should be done. Importantly, REVO accommodates innovation and accidental discovery through a decision-making (feedback) cycle. The intent of the REVO decision-making cycle is to ensure a research program is responsive to its operational environment by enabling discovery and harnessing those discoveries that matter. At all times, REVO-related information is managed in an integrated manner, even if selected information is not connected or is contradictory.

The specific objective of this article is to provide a concise but comprehensive review of the REVO method using an example from the cybersecurity domain to demonstrate the utility of the approach. We plan to further refine the approach as our understanding deepens and our experience grows.

In the first section of this article, we describe how strategic research contexts and research-requirement statements are used to articulate what needs to be done. In the second section, we describe how research-activity descriptions are used to track what is being done (including when providing information for Statistics Canada's Federal Scientific Expenditure and Personnel [FSEP; tinyurl.com/l9j2p22] survey). In the third section, we describe the lifecycle of the research program and explain how key performance indicators and a decision-making cycle are used when assessing the overall progress with R&ED. The cybersecurity example that is used throughout this article to illustrate the REVO approach is directly linked to the research focus area "Investigate practical enterprise-level metrics" described in the companion article by Craigen, Walsh, and Whyte (2013; timreview.ca/article/704).

Articulating What Needs To Be Done

In this section, we summarize the components of REVO used for describing the key challenges that drive our R&ED program. The first sub-section presents the notion of strategic research contexts, which we view to compose the breadth of our problem space. The second sub-section presents the notion of research-requirement statements, which are structured expressions of specific problems, and which we view to compose the depth of our problem space. We have found it useful to be able to: i) concisely summarize the challenge space overall; ii) separately describe specific problems in a fine grained and focused way; and iii) link these tightly scoped statements to a broader scope. When analyzing the link(s) that may exist from a specific research-requirement statement to one or more strategic research contexts, it becomes clear why the requirement is relevant with respect to the overall research program. When analyzing the link(s) that may exist from a particular strategic research context to one or more research-requirement statements, it becomes clear how well that aspect of the problem domain is understood and what specific research-related activities should be pursued.

Strategic research contexts

Strategic research contexts (SRCs) compose the breadth of our cybersecurity problem space. SRCs further explicate portions of our cybersecurity challenges and therefore inform the coverage of research requirements and alignment of research programs and activities. Based on our experiences with cybersecurity and discussions with other stakeholders in the domain, we have identified 19 SRCs that provide structure to the problem space (Box 1). We believe that research advances in these contexts will help achieve a stable and resilient information technology infrastructure for Canada.

Research-requirement statements

In the REVO process, the information technology department of the company's cybersecurity manager specifies what is needed using a template for a research-requirement statement. A simplified example of a completed research-requirement statement is provided in Appendix A (tinyurl.com/n6vkm82) using data from a fictitious enterprise. The example focuses upon a requirement for well-founded security measures and metrics. Though simplified here, the topic is a valid cybersecurity research requirement.

A research-requirement statement consists of 10 sections:

Section 1. Identification/Criticality: consists of basic information including date, identification number, a title, point of contact, and group, urgency and importance. In our example, we note that the requirement is urgent and of high importance to the enterprise (from the perspective of the business line).

Managing Cybersecurity Research and Experimental Development: REVO Approach

Dan Craigen, Drew Vandeth, and D'Arcy Walsh

Box 1. Strategic research contexts for cybersecurity

SRC1 – Mission Management
Comprises policy, priority, resource, and risk management in support of optimizing mission effectiveness.

SRC2 – Computational Platforms
Comprises forms of computation systems as a means of implementing particular types of algorithms, satisfying operational constraints, managing computation resources, and includes the application of specific approaches of interest.

SRC3 – Autonomous and Adaptive Systems
These are systems that are designed to respond automatically and without intervention to some range of environmental or operating conditions. Communities of heterogeneous or homogeneous systems can interact cooperatively among themselves and with the environment, or possibly dynamically reconfigure themselves, to meet a set of common mission goals.

SRC4 – Human–Computer Interaction
Includes all logical and physical forms of interface between humans and computers. Varieties of interaction are needed to suit the types of complex and voluminous mission information that humans must interpret and manipulate.

SRC5 – Sensor Architecture
Situational awareness and intelligent network management for cybersecurity require sensor architectures that identify host-based and network-based events and may enrich both situational awareness and network management by performing host-based, network-based, or combined analytics. Such architectures may require complex command and control capabilities.

SRC6 – Database Systems
Data must be represented, stored, manipulated, filtered, and retrieved to suit particular mission purposes and conditions.

SRC7 – Secure System Architecture
A set of system attributes, described in design artifacts, that specify how they relate to the overall IT architecture. These controls serve the purpose of maintaining the system's quality attributes, among them confidentiality, integrity, availability, accountability, and assurance.

SRC8 – Cryptanalysis
Used to characterize systems and to characterize vulnerabilities in encryption methods to access encrypted information. Methods can be mathematical, protocol based, or based on the physical-system implementation.

SRC9 – Computer Network Analysis
Used to characterize networks and to characterize vulnerabilities in networks that may be used to disrupt intended network functionality. A broad class of methods, drawing upon interdisciplinary techniques, must be understood for protecting modern cybersystems.

SRC10 – Trusted Computing
Provides the means to create trustworthy computational systems in environments that cross security domains. Trusted computing includes evaluation of expected software and hardware function and acceptable deployed risk of vulnerability; it also includes the development of methods of detecting, mitigating, and preventing compromises of system security. Trusted computing depends on techniques for constructing systems that are inherently secure at some level.

SRC11 – Computer Network Defence
Develop techniques to detect, assess, and respond to cyberintrusions of networks and systems. Computer network defence is informed by elements such as sensor architectures, computer network analysis, security measures and metrics, and knowledge discovery.

SRC12 – Security Measures and Metrics
Provide a quantitative and objective basis for security assurance, with the main uses being for strategic support, quality assurance, and tactical oversight. Metrics can be applied to measure the maturity of security processes or of the security posture.

SRC13 – Secure Communications
The creation of systems that allow two parties to communicate in a way that is insusceptible to eavesdropping or interception.

Continued on next page...

Box 1. (*continued*) Strategic research contexts for cybersecurity

SRC14 – Knowledge Discovery
An interdisciplinary field focusing on methodologies for extracting useful knowledge from data, drawing upon statistics, databases, pattern recognition, machine learning, data visualization, optimization, and high-performance computing. Knowledge discovery includes the efficient preparation, display, summarization, search, and filtering of complex data sets.

SRC15 – Distributed Computational Space
Development of analysis, filtering, retrieval, and other processing techniques for operating in a distributed computational environment either by their inherently distributed nature or distributed by constraint.

SRC16 – Advancing Analytics
Techniques for advanced logical analysis of data and human behaviour for a mission purpose, supporting an analytic process to make it more efficient, to make it more effective, to manage it, and to automate it. Data may be of large scale and from disparate sources, requiring different methodologies for understanding it.

SRC17 – Systems Engineering
The robust approach to the design, creation, and operations of systems. Systems engineering includes the specifying of system goals as well as articulating design concepts, tradeoffs, implementation, and verification.

SRC18 – Material Science
The application of advanced materials and fabrication techniques to enable other technologies and to support mission systems.

SRC19 – Cyber–Physical Systems (CPS)
Those systems in which there is a strong connection between computational (cyber) and physical elements. Much of our critical Infrastructure depends upon cyber–physical systems. Human-in-the-loop cyber–physical systems are those systems that consist of a human, an embedded system, and the physical environment. Human-in-the-loop systems can restore fundamental autonomy for functionally weakened individuals. A robust cyber-security framework will encourage deployment of cyber–physical systems, including human-in-the-loop systems.

Section 2. Stakeholders: the key operational stakeholders are identified. Normally, an operational stakeholder, a technical stakeholder and a subject-matter expert. In our example, we identified "Information Operations" and "Enterprise Security" as enterprise stakeholders and Mike Smith and John Doe as two subject matter experts.

Section 3. Business Description: here, the business motivations for the research requirement are captured. We give three example motivations in Appendix A, including the observation that it has become increasingly difficult to choose amongst security options because the benefits, costs, and tradeoffs are poorly understood.

Section 4. Research Requirement: the specifics of the research requirement are captured in this section, including technical challenges and proposed solutions or approaches. In our example, we observe that advancing the state of scientifically sound security measures and metrics would greatly aid the design, implementation, and operation of secure information systems.

Section 5. Success/Completion Criteria: often overlooked is a statement of how one knows that a research requirement has been resolved. Security measures and metrics are sufficiently immature that it is difficult to fully identify success. However, we would hope to ensure that: i) the security posture is continuously monitored; ii) the measurements meaningfully reflected security posture; and iii) both manual and automated responses to appropriate classes of threats are suitably informed.

Section 6. Category Impact: this refers to potential impact either to the enterprise, partners, or general enterprise research capabilities. The impact is low, medium, or high. In our example, the research is expected to have a high impact on the enterprise's operational capabilities.

Section 7. Description of Impact: reasons are provided for claims of a particular impact. For the example, we believe that enhanced understanding of the IT infrastructure will identify attack vectors and vulnerabilities and will better inform what system data is required.

Section 8. Relationship to Strategic Research Contexts: the depth of the requirement is tied to the breadth of the strategic research contexts. In this case, the link is particularly simple because the research requirement maps to SRC12: Security Measures and Metrics.

Section 9. Partnerships: researchers that we could leverage or partner with in advancing the requirement. In our example, we point to three institutions in the United States that are working in the area of "science of security" and have identified security measures and metrics as a hard challenge.

Section 10. Notes: a free-form section for any addition information the business line wishes to provide.

Assessing the research requirements

Based upon information at hand and guidance from the business lines, the requirements are categorized as advancing enterprise operational capabilities, advancing partner operational capabilities, or advancing enterprise research capabilities. Within these categories, research requirements are then tiered into three levels of enterprise criticality, with Tier I being the most critical.

Assessing the tier of a research requirement is based on the following five criteria:

1. Coverage of the strategic research contexts

2. Importance or impact within its category

3. Originator criticality specification (intra-research-requirement statement)

4. Other research-requirement statements (inter-research-requirement statement)

5. Retrospective information (heuristics, lessons learned)

Based on the resulting tier, the following requirements/focus areas are recommended:

1. Tier I requirements: should address critical internal-research issues; should be specified in a manner that is actionable by internal research capacity; are usually more granular and narrower in scope; and should be owned by a business-line research effort.

2. Tier II requirements: should supplement or augment internal research issues; should be specified in a manner that is actionable by internal research capacity, but primarily to drive the investigations of external research capacity to address the broader context; are usually more coarse grained and broader in scope; and should be owned by a business-line research effort.

3. Tier III requirements/focus areas: should be identified to drive predictive analysis investigations to supplement Tier I and Tier II investigations. (A research focus area, for example simulation techniques, identifies a general technical area of potential interest.)

Based on our experiences, we identified the following three options for applying appropriate resources:

1. Utilize internal research capacity

2. Form and manage external research relationships

3. Use predictive-analysis methods and techniques

In general, we recommend that, because of the breadth and depth of the problem space and the often-limited internal research capacity of the organization, such capacity should be focused on Tier I problems. Hence, for a Tier I requirement, we suggest that an optimal combination of the three resources be applied. For Tier II, a combination of external research relationships and predictive-analysis methods and techniques is optimal. For Tier III, predictive-analysis methods and techniques are appropriate. Optimization of resources should be determined on a case-by-case basis depending upon, for instance, organizational capacity, partnerships and funding.

Tracking What Is Being Done

In this section, we summarize the components of REVO that are used for specifying and tracking specific research endeavours that are planned, that are in progress, or that have been completed. In the following sub-section, we present the notion of research-activity descriptions and describe how they are aggregated to respond to the annual Federal Science Expenditures and Personnel (FSEP; tinyurl.com/l9j2p22) survey.

Research-activity descriptions

Research-activity descriptions are structured descriptions of specific internal or external investigations that have established resourcing levels in terms of expenditures or personnel. A specific research-activity description is linked to one or more research-requirement statements. When analyzing the link(s) that may exist from a specific research requirement statement to one or more research-activity descriptions, it becomes clear what level of effort is required, including an indication of what degree of progress may be anticipated. When analyzing the link(s) that may exist from a specific research-activity description to one or more research-requirement statements, it becomes clear what impact the specific activity may have, or not have, depending upon the outcome of the particular investigation.

A simplified example of a completed research-activity description pertaining to security measures and metrics is provided in Appendix B (tinyurl.com/kzee5mj). A research-activity description consists of two parts:

Part 1: General information: consists of basic information such as fiscal year, point of contacts, and linkage to strategic research contexts.

Part 2: Research-activity information: consists of eight sections that provide particulars of the project. These sections are:

Section A. Project identification: further elaborates basic project information including a statement of the purpose of their work and the kind of work (experimental development or advancement of scientific knowledge). Depending upon the response, either Section B or Section C will be completed.

Section B. Experimental development: determines what technological advancements are being targeted, what technological obstacles exist, and what work has been directed at overcoming the obstacles. In our ongoing example, we discuss collecting known measures and metrics into a single compendium and then experimenting using an in-house enterprise laboratory. We state that the main obstacle is the identification of measurements and metrics of suitable quality. The uncertainty of the work is noted by the explicit statement that the work is sufficiently immature that it is unclear how the technical obstacles will be overcome.

Section C. Basic or applied research: though this section was not completed in our example, it determines what scientific knowledge is being progressed, what work is to be performed, and how it contributed to the scientific knowledge.

Section D. Additional project information: identifies the collateral developed by the project, such as planning documents, resource allocation, notebooks, and contracts.

Section E. Intramural expenditures: essentially captures internal expenditures.

Section F. Extramural expenditures: essentially captures external expenditures.

Section G. Personnel: determines how many individuals (measured as full-time equivalents) worked on the project or supported the project.

Section H. Sources of funds: determines where the funds come from.

Note that Sections E through H of the research-activity description provide the project's specific financial and staffing figures for the FSEP survey. The organization's response to FSEP will aggregate the figures from all of their R&ED projects.

Research Program Lifecycle

In this section, we summarize the components of REVO used to manage the research program lifecycle as a whole to ensure R&ED efforts result in required operational capability in an efficient and effective manner. We first present the key performance indicators (KPIs; tinyurl.com/ltsjzja) that are used to set targets (through establishing thresholds) and assess progress. Then, we present the decision-making cycle that is used to (re)align the research program when adapting to changing business, technical, legal, and other drivers or constraints.

Key performance indicators

The following four KPIs enable the full lifecycle of our R&ED program to be continually (re)assessed with respect to established and emerging research priorities. When other components of REVO change, the KPIs are recomputed. We believe they are useful high-level indicators that can apply to any domain under investigation.

KPI1. Alignment of research-requirement statements and strategic research contexts: provides a top-level indication of how well the breadth of the problem space is covered by research-requirement statements. This KPI is computed by setting thresholds for each strategic research context relating to the percentage of research requirements that are expected to be linked to that context. This dashboard-like indicator "goes red" when one or more strategic research contexts lack actionable problem statements.

KPI2. Balance of internal and external investigations: provides a top-level indication of how well balanced internal and external investigations are given internal resources. Ideally, internal capacity should only be used to address Tier I research requirements. This dashboard-like indicator "goes red" when internal resources are directed at research requirements that should ideally be addressed by external resources.

KPI3. Distribution of expenditures and personnel resourcing levels: provides a top-level indication of the distribution of research-related resources with respect to Tier I, Tier II, and Tier III research requirements. This KPI is computed by setting thresholds for each tier relating to the percentage of resources that are expected to be allocated to that tier. This dashboard-like indicator "goes red" when a particular tier is underfunded to the benefit of another tier.

KPI4. Assessment of progress of activities: provides a top-level indication of the progress of the research program as a whole with respect to the completion criteria that were specified for each research-requirement statement. This KPI is an aggregated result of assessments made by subject-matter experts about whether limited progress (0), high-potential progress (1), or definite progress (2) is being made for each active research initiative. This dashboard-like indicator "goes red" when the research program is not producing results effectively or efficiently.

The decision-making cycle

In this sub-section, we describe the high-level decision-making cycle that is used to keep the research program as a whole responsive to changing operational priorities. The research-program executive sets direction by validating the strategic research contexts and setting the thresholds that are used to compute KPIs. Subject-matter experts are responsible for articulating research-requirement statements and assessing progress of particular investigations. Research managers are responsible for tracking research activities that are part of their portfolio. When one or more indicators "turn red", decisions are taken to turn the indicator(s) back to green. Depending upon the indicator, this may mean:

- readjusting the representation of the strategic research contexts

- adding, deleting, or refining research-requirement statements

- changing the mapping between strategic research contexts and research-requirement statements

- adding, deleting, or refining research-activity descriptions

- changing the mapping between research-requirement statements and research-activity descriptions

- adjusting resource levels with respect to Tier I, Tier II, and Tier III research requirements

- adjusting thresholds

Conclusion

In this article, we have presented a high-level description of REVO using a specific cybersecurity requirement and activity description that are linked to the breadth of the cryptologic problem space. The examples are intended to illuminate the key artifacts that we have found give REVO its power as a practical and flexible systematic approach for managing R&ED. Due to time and space limitations, we have not been able to provide complete examples nor to report upon refinements specific to our work context. As our understanding deepens and our experience grows, we plan to publish more in-depth articles about how REVO enables us to address the cybersecurity context, principles, and focus areas described in the companion article by Craigen, Walsh, and Whyte (2013; timreview.ca/article/704).

We conclude by making comments about: i) the use of specific methods to address specific problems and ii) the use of a general methodology for a unified response to a large and complex R&ED challenge.

Use specific methods to address specific problems

Based upon our academic work and our ongoing investigations in the workplace, we understand that it is important to: i) have a clear and well-scoped understanding of the specific problem under investigation and ii) be explicit about the particular methodological approach that will be applied when pursuing an investigation. The methods applied should be "as strong as possible" in the sense that some methods may be more applicable than other methods, depending on the problem.

We advocate always specifying the particular methodological approach that will be adopted, coupled with the description of the specific problem of concern. As an investigation proceeds, the methodology should be evaluated along with reporting any research results with respect to the problem itself.

Use a general methodology for a unified response to the challenge
We also recognize the need for applying a general methodology to facilitate a unified response to the challenge overall. In our view, this methodology needs to be "strong enough". A unifying method must balance rigour with flexibility. The general method must be rigorous enough to provide traceability to top-down and bottom-up objectives, including the quantification of performance metrics for R&ED. The method must also be flexible enough to accommodate the potentially highly divergent approaches that could be trialed on a problem-by-problem basis.

The general methodology should be well-informed by the definitions and methodologies pertaining to R&ED as espoused by the OECD's Frascati Manual (tinyurl.com/kq44wqx) for measuring scientific and technological activities.

Appendices

A. Example Research-Requirement Statement
Available online at: tinyurl.com/n6vkm82

B. Example Research-Activity Description
Available online at: tinyurl.com/kzee5mj

About the Authors

Dan Craigen is a Science Advisor at the Communications Security Establishment Canada (CSEC). Previously, he was President of ORA Canada, a company that focused on High Assurance/Formal Methods and distributed its technology to over 60 countries. His research interests include formal methods, the science of cybersecurity, and technology transfer. He was the chair of two NATO research task groups pertaining to validation, verification, and certification of embedded systems and high-assurance technologies. He received his BScH in Math and his MSc in Math from Carleton University in Ottawa, Canada.

Drew Vandeth is the Senior Research Strategist for the National Security Community and a Senior Researcher at IBM Systems Research. He is the founder of the Tutte Institute for Mathematics and Computing (TIMC) and was its first Deputy Director. His research interests include theoretical and computational number theory, contextual and cognitive computing, high performance computing architectures, autonomic and autonomous analytical systems, and research management. Dr. Vandeth holds a PhD in Number Theory from Macquarie University in Sydney, Australia, an MMath in Number Theory from the University of Waterloo, Canada, and a BMath (Hons) in Pure Mathematics, also from the University of Waterloo.

D'Arcy Walsh is a Science Advisor at the Communications Security Establishment Canada (CSEC). His research interests include software-engineering methods and techniques that support the development and deployment of dynamic systems, including dynamic languages, dynamic configuration, context-aware systems, and autonomic and autonomous systems. He received his BAH from Queen's University in Kingston, Canada, and he received his BCS, his MCS, and his PhD in Computer Science from Carleton University in Ottawa, Canada.

Citation: Craigen, D., D. Vandeth, and D. Walsh. 2013. Managing Cybersecurity Research and Experimental Development. *Technology Innovation Management Review.* July 2013: 34–41.

Keywords: cybersecurity, research, experimental development, performance indicators, strategic research contexts, research-requirement statements, research-activity descriptions, research program lifecycle

Security Challenges in Smart-Grid Metering and Control Systems

Xinxin Fan and Guang Gong

> *"As we modernize the nation's electric infrastructure to make it smarter, more efficient, and more capable, we need to make it more secure from end to end."*
>
> Gary Locke
> U.S. Ambassador to China
> and Former Secretary of Commerce

The smart grid is a next-generation power system that is increasingly attracting the attention of government, industry, and academia. It is an upgraded electricity network that depends on two-way digital communications between supplier and consumer that in turn give support to intelligent metering and monitoring systems. Considering that energy utilities play an increasingly important role in our daily life, smart-grid technology introduces new security challenges that must be addressed. Deploying a smart grid without adequate security might result in serious consequences such as grid instability, utility fraud, and loss of user information and energy-consumption data. Due to the heterogeneous communication architecture of smart grids, it is quite a challenge to design sophisticated and robust security mechanisms that can be easily deployed to protect communications among different layers of the smart grid-infrastructure. In this article, we focus on the communication-security aspect of a smart-grid metering and control system from the perspective of cryptographic techniques, and we discuss different mechanisms to enhance cybersecurity of the emerging smart grid. We aim to provide a comprehensive vulnerability analysis as well as novel insights on the cybersecurity of a smart grid.

Introduction

The term "smart grid" generally refers to a next-generation power grid in which the generation, transmission, distribution, and management of electricity are upgraded and automated by incorporating advanced computing and communication technologies for improving the efficiency, reliability, economics, and safety of the grid. Loosely speaking, a smart grid is composed of a power grid and a two-way communication network for information retrieval and management. When compared to legacy and closed power-control systems, the smart grid is envisioned to establish a scalable, pervasive, and interactive communication infrastructure with new energy-management and demand-response capabilities. During the past few years, smart-grid metering and control systems have been widely deployed throughout the world. According to a new Navigant Research report (2013; tinyurl.com/m3qm7xx), the global market potential for smart-grid equipment manufacturers and solution providers will nearly double by 2020, reaching $73 billion in annual revenue and $461 billion in cumulative profit.

A smart grid brings great performance benefit to the power industry and enables end users to optimize their power consumption; however, the heavy dependence on communication networks has made smart grids vulnerable to a wide range of cyberspace threats. For example, it has been shown that security breaches in smart grids can result in a variety of serious consequences, from blackouts and physical damage of infrastructure to the leakage of customer information. Considering the vast scale and complex architecture of a smart grid, it is not difficult to understand that the vulnerabilities associated with the smart-grid communication system may also be enormous. Those security vulnerabilities need to be properly addressed to ensure that smart grids are not only secure and function correctly, but that they also maximize their adoption and successfully fulfill the promise of smart-grid investment.

Although most of the architectures, frameworks, and roadmaps for smart grids have already been defined by the governments, industry, and academia, there are still many important security and privacy issues in smart-grid communications. These issues are now considered by governments and industry to be one of the highest priorities for smart-grid design, and they must be resolved before smart grids can be operationally ready for the market. In this article, we will present the high-level architecture of a smart-grid metering and control system, and we will describe typical cyberspace attacks on smart-grid communications. We also will summarize the security requirements, review some existing solutions, and highlight several important directions along this emerging research line.

The remainder of this article is organized as follows. First, we present the fundamental architecture and functionalities of a smart-grid metering and control system. Next, we focus on the security requirements for smart-grid communications, followed by a survey of current efforts made by the industry and academia to secure the smart-grid networks and devices. Finally, we propose several research areas and directions in smart-grid security and draw some conclusions.

Architecture

A typical smart-grid metering and control system, as illustrated in Figure 1, consists of a collection of meters/sensors and controllers/actuators that communicate with a substation/data-concentrator, a consumer or technician, and various third-party entities. The communication among different network entities is realized by high-speed wired or wireless links or a combination thereof. A smart-grid metering and control system has a layered network structure through which it collects data and controls the delivery of electricity.

The main functionalities of each component in a smart-grid metering and control system are as follows:

1. **Utility company:** connects to the substation network through the wide area network (WAN) interface and the communication channel might be Wi-Fi, satellite, 4G-LTE, Wi-Max, etc. The utility company is responsible for processing alarms and alerts, managing the meter data, and generating bills. Moreover, it may also provide a web portal that allows customers to view their monthly energy consumption and bills.

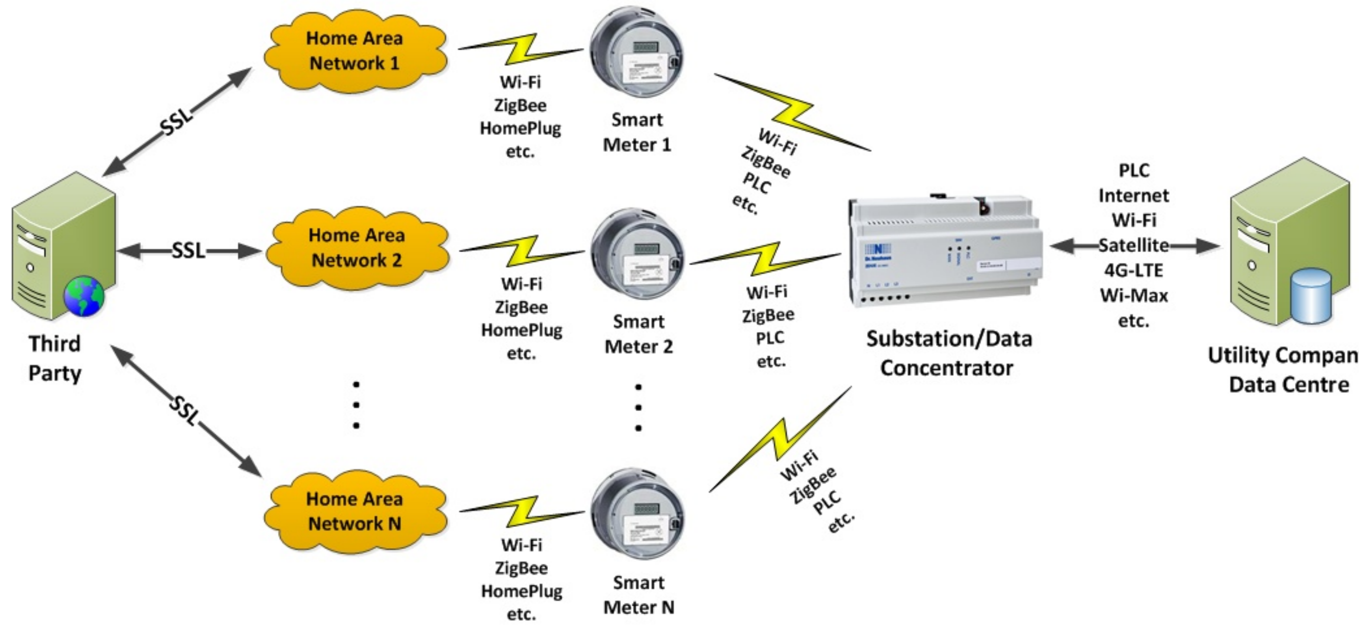
Home Area Network 1
Home Area Network 2
Home Area Network N
SSL
SSL
SSL
Third Party
Wi-Fi ZigBee HomePlug etc.
Wi-Fi ZigBee HomePlug etc.
Wi-Fi ZigBee HomePlug etc.
Smart Meter 1
Smart Meter 2
Smart Meter N
Wi-Fi ZigBee PLC etc.
Wi-Fi ZigBee PLC etc.
Wi-Fi ZigBee PLC etc.
Substation/Data Concentrator
PLC Internet Wi-Fi Satellite 4G-LTE Wi-Max etc.
Utility Company Data Centre

Figure 1. Architecture of a typical Smart-grid metering and control system

2. **Substation/data-concentrator network:** consists of a number of smart meters in a certain area as well as a data collector. The connection between smart meters and the data collector might through Wi-Fi, ZigBee, power line carrier (PLC), etc. Typically, the smart meters form a wireless mesh network and forward the meter readings to the data collector through multi-hop communications. The data collector then transmits the accumulated data to the utility company.

3. **Home area network (HAN):** provides the consumer access points to control and monitor the real-time power consumption. The HAN contains a home gateway that receives the power-consumption data from the smart meter and displays it on householder's devices (e.g., laptop, tablet, smartphone). Furthermore, the home gateway may send the power consumption data to a third party for other value-added services (e.g., efficiency advice, supplier selection). The HAN also includes a controller that enables householders to remotely control the status of their home appliances.

4. **Smart meter**: is composed of a microcontroller, a metrology board, and a communication board. Under the control of the microcontroller, the metrology board measures the real-time power consumption, and the meter data is transmitted to both the substation network as well as the home area network through the communication board. The connection between the smart meter and home appliances may be through Wi-Fi, ZigBee, Ethernet, HomePlug, Wireless M-Bus, etc. The smart meter may also contain a disconnection function that (if enabled) allows utility companies or customers to remotely connect or disconnect the home appliances and services.

5. **Third party:** relies on accurate meter readings to provide value-added services for householders, including power efficiency advice, supplier selection, etc. Those services will help householders to manage their power usage in a cost-effective way.

Requirements

The conventional power grid is composed of dedicated power devices that form closed networks with reliable and predicable communication links. In contrast, a smart-grid metering and control system relies on advanced wired and wireless communication networks, thereby inheriting all of the weaknesses and potential cyberspace vulnerabilities of general communication networks. The smart-grid metering and control system is becoming an increasingly common target for cyberspace attacks, and strong and robust security mechanisms are paramount for the prevention of financial fraud, environmental accidents, and a host of other potentially disastrous incidents. In this section, we discuss the major security concerns and requirements for smart-grid metering and control systems.

Efforts from standards bodies and organizations
A number of organizations have been actively working on the development of smart grid security requirements, as illustrated in Box 1. Among existing smart-grid standardization efforts, the NIST Framework and Roadmap for Smart Grid Interoperability Standards and its Interagency Report, “Guidelines for Smart Grid Cyber Security” (NIST IR 7628; tinyurl.com/yb6jpuw), represent the most comprehensive coverage of cyberspace security requirements in the smart grid.

All standards bodies consistently specify three high-level smart-grid security objectives: availability, integrity, and confidentiality. However, even though the standards bodies define the security requirements based on a fairly comprehensive set of use cases in the power industry, there is still a considerable gap between understanding the security requirements in the standards and applying them to design a secure-

Box 1. Examples of organizations working on smart-grid requirements

- Electric Power Research Institute (EPRI; epri.com)
- International Society of Automation (ISA; isa.org)
- IEEE 1402-2000 (tinyurl.com/ox786r8)
- International Electrotechnical Commission (IEC; iec.ch)
- National Energy Board (NEB, Canada; neb-one.gc.ca)
- North American Electrical Reliability Corporation-Critical Infrastructure Protection (NERC-CIP; nerc.com)
- National Institute of Standards and Technology (NIST; nist.gov)

smart grid metering and control system. It is extremely important for designers and practitioners of smart grids to gain deep understanding about a wide range of malicious attacks to the smart grid, as detailed below.

Availability
Availability refers to ensuring timely and reliable access to information, which is the primary security goal of a smart-grid metering and control system. Malicious attacks targeting availability can be considered as denial-of-service attacks (tinyurl.com/jzn67), which intend to delay, block, or even corrupt the communication in the system. In particular, due to the extensive adoption of wireless communication technologies in the smart grid, a jamming attack (tinyurl.com/km9sd9) that fills the wireless medium with noise signals has become the most typical form of physical-layer attack. The jamming attack is able to defer the transmission of messages and to distort the transmitted data signal. As a result, the legitimate receiver cannot recover messages out of the damaged data packets. Jamming attacks are more relevant and serious in the smart grid than other than other networking systems, because the smart grid involves essential resources for people's everyday lives. On the other hand, many man-in-the-middle attacks (tinyurl.com/fco32) can be launched only when the full or partial communication channels can be jammed. Examples include jamming then inserting false location information and jamming then delaying the transmission. Because the network traffic in the smart grid is generally time-critical, it is crucial to evaluate the impact of denial-of-service attacks and to design efficient and effective countermeasures to such attacks.

Integrity
Integrity refers to preventing or detecting the modification or destruction of information by unauthorized persons or systems. Malicious attacks targeting the integrity of a smart grid attempt to stealthily manipulate critical data such as meter readings, billing information, or control commands. Recent research (Liu et al., 2011; tinyurl.com/kzaxzdy) has demonstrated that a new class of attacks, called false data-injection attacks, are highly viable against the state estimation in electrical power grids. Based on the assumption that an attacker has compromised one or several smart meters and is able to access the current power-system configuration information, such attacks can successfully inject arbitrary bogus data into the monitoring centre, and at the same time, pass the data-integrity checking used in current state-estimate processes. Integrity protection can be achieved by authentication, certification, and attestation. More specifically, the smart devices and substation must authenticate each other's identity to thwart impersonation. Data certification of a message prevents modification of data during transmission. Data authentication with non-repudiation goes beyond certification by preventing the sender from claiming that it did not send the data. Substations use attestation to confirm that the memory contents (code and data) on a smart device have not been modified. The security services related to integrity are usually implemented using public-key cryptography, which requires a trusted third party that hosts a key-management service.

Confidentiality
Confidentiality refers to protecting personal privacy and proprietary information from unauthorized access. Malicious attacks targeting confidentiality aim at obtaining desirable information (e.g., power usage, customer's account information) through eavesdropping on communication channels in a smart-grid metering and control system. Although such attacks have negligible effects on the operation of the system, the transmission of fine-grained consumption data by smart meters has raised concerns about privacy. Research (Quinn, 2009; tinyurl.com/pc2st2e) has shown that the consumption data collected by smart meters reflects the use of all electric appliances by inhabitants in a household over time, and it allows criminals to make inferences about the behaviours, activities, or preferences of those inhabitants. Those privacy issues need to be addressed appropriately to reduce customers' fears about potential leakages of their information. Some best practices relating to privacy have been proposed for the design of smart grids (Cavoukian, 2010; tinyurl.com/27r43ds). An emerging trend is for the smart meters to aggregate usage data for billing purposes and support load-balancing and other monitoring functions through peer-to-peer protocols that preserve the consumer's privacy.

Current Approaches

Based on the security guidelines specified by the NIST and other standards bodies, both industry and academia have made efforts to address the challenging security issues in smart-grid metering and control systems by employing various cryptographic techniques. Here, we give an overview of several existing cyber-security solutions proposed by industry and academia for smart-grid communications.

Security Challenges in Smart-Grid Metering and Control Systems

Xinxin Fan and Guang Gong

Cybersecurity solutions from industry

In 2007 a large stakeholder community was assembled by the ZigBee Alliance to address the security issues in the smart grid; this community developed what is known as the ZigBee Smart Energy Profile (SEP; tinyurl.com/kjfl96m). The ZigBee SEP has been widely adopted as the communication infrastructure in home area networks. Regarding to the security, the ZigBee SEP specifies that each smart meter should be equipped with an Elliptic Curve Qu-Vanstone (ECQV; tinyurl.com/7v4f36a) implicit certificate before deployment. The ECQV certificate is much smaller than a traditional X.509 certificate (tinyurl.com/8e3zr), and it binds a meter's MAC address and manufacture identifier to an ECC key pair (tinyurl.com/egz7y). Although the ECQV certificate issuance has been addressed (Certicom; tinyurl.com/mbug9b), the certificate renewal and revocation processes are not defined in the ZigBee SEP.

For supervisory control and data acquisition (SCADA; tinyurl.com/jcrlz) systems, NIST (2010; tinyurl.com/mfrn42j) suggests AES, SHA-1, and RSA, and IEC 62351 (tinyurl.com/29tm8ll) specifies RSA-1024. However, it is now known that RSA is a poor choice for SCADA networks because of the high computation cost of RSA encryption and the limited computing power of SCADA devices. The Standards Council of Canada (tinyurl.com/m4najzg) and the European Union (tinyurl.com/kvmnswk) also define cybersecurity requirements for smart grids, but do not specify a suite of cryptographic algorithms to meet the requirements, except that the Standards Council of Canada specifies that SHA be used as the secure hash function. It remains an open research problem to find a set of cryptographic algorithms that provide the right combination of security and implementability for the smart-grid metering and control system.

Besides industry alliances and standards bodies, there are a number of manufacturers of smart devices for SCADA networks and meters for smart grids. Implementation details for these devices are generally considered proprietary information, but a few generalizations can be made. The cryptographic algorithms are implemented in software on a low-power 16-bit microprocessor. RSA-1024 or ECC-256/384 is used for public-key services. Symmetric key services use AES-128 or AES-256. Some devices use spread-spectrum modulation. Most smart-device manufacturers implement the security services themselves. A few companies have a hardware security module (HSM; tinyurl.com/7r9v6rv) or similar product that is independent of a specific smart device. SafeNet's PKI HSM (tinyurl.com/k6mbobz) provides public key cryptography with RSA-1024 and ECC-256/384, and symmetric-key cryptography with AES-256 to perform attestation, key management, encryption/decryption, and billing. GE Digital Energy (tinyurl.com/lsjyqsl) makes a family of wireless routers with AES-128 designed to connect to smart meters and controllers. Within Canada, Tofino Security's Industrial Security Solution (tofinosecurity.com) is a server-side software program combined with security devices that act as wired access points with encryption for meters and actuators. Bentek Systems' SCADALink SMX900 (tinyurl.com/mrj74mb) is a modular wireless remote-terminal-unit/modem that supports spread-spectrum communication, but does not appear to have any facilities for encryption, authentication, etc.

Cybersecurity solutions from academia

A critical component of smart grid security is key management, which will ensure the confidentiality, authenticity, and integrity of devices and communications within the grid. Most previous research focused on designing cryptographic protocols to provide certain security functionalities.

Efficient implementations of encryption schemes are essential for providing confidentiality in a smart grid. An experimental study about the performance of a symmetric-key cipher (i.e., DES-CBC) and a public-key cipher (i.e., RSA) on an intelligent electronic device (IED) called TS7250 has been conducted (Wang and Lu, 2013; tinyurl.com/mlzypxp), where the IED is used for sending the transformer status and receiving commands from the control centre. These experimental results show that the computational ability of an IED becomes a bottleneck for the delay performance when performing asymmetric-key cryptography. These authors also suggested that a symmetric-key approach is more suitable for real-time IED communications in power distribution and transmission systems.

Authentication is crucial to protect the integrity of data and devices in the smart grid. Due to the limited computational capabilities of devices, stringent timing requirements, and high data-sampling rates in the smart grid, traditional authentication schemes might not be applicable. Moreover, besides supporting basic data and device authentication, multicast authentication is another desirable feature due to the multicast nature of the smart-grid communication. A number of authentication schemes have been proposed in the literature for smart grids. Szilagyi and Koopman (2009: tinyurl.com/k8pwh46 and 2010: tinyurl.com/l93xwjs) proposed flexible and low-cost multicast authentication schemes for embedded control systems. The basic idea is to verify truncated message authentication codes (MACs) across multiple packets, thereby achieving a good tradeoff among authentication cost, delay performance, and tolerance to attacks. Wang and colleagues (2009; tinyurl.com/qxvb49v) proposed a fast multicast authentication scheme for time-critical messages in the smart grid. Their scheme is based on an efficient variant of a one-time signature (OTS) scheme. Although the proposed scheme is efficient in terms of computation, the public key size in an OTS-based scheme is quite large (i.e., on the order of 10KB). Hence, both communication and storage overhead are significant in this case. Lu and colleagues (2012; tinyurl.com/m3xfj5g) conducted an empirical study for a few data-origin authentication schemes in substation automation systems (SAS). These authors compared the performance of RSA, MAC, and OTS on a small-scale SAS prototype and concluded that the existing authentication schemes cannot be applied directly into the SAS due to insufficient performance considerations in response to application constraints.

The heterogeneous communication architecture of the smart grid has made the key management particularly challenging, and it is not practical to design a universal key-management scheme for the entire smart grid. The simplest way is to use a single key shared by all the meters in the smart grid. However, this solution will cause the single point of failure due to the lack of a tamper-proof module in smart meters. Beaver and colleagues (2002; tinyurl.com/qcgsgth) proposed an elementary key-establishment scheme called SKE for SCADA systems. Whereas the master-slave communications are secured by symmetric-key schemes, the peer-to-peer communications are protected by public-key schemes. However, the scheme proposed by these authors does not support efficient multicast and broadcast authentication in the smart grid. Dawson and colleagues (2006; tinyurl.com/lkkoxgb) proposed SKMA, a key management scheme for SCADA systems. These authors introduced a key-distribution centre (KDC) and each node maintains two types of long-term keys: node-to-KDC and node-to-node. A session key in SKMA is generated using the node-to-node key. Unfortunately, SKMA does not consider issues of multicast, key update, and revocation. Choi and colleagues described ASKMA (2009; tinyurl.com/mooapta) and ASKMA+ (2010; tinyurl.com/ml2kvqm) for key management in SCADA systems, respectively. Both schemes are designed based on the usage of a logical key hierarchy (LKH), which is able to achieve efficient key management among all nodes. In particular, ASKMA supports both multicast and broadcast authentication and the performance has been further improved in ASKMA+.

Although many encryption, authentication, and key-management schemes have been proposed, their performance does not seem to fulfill the stringent timing requirements of the smart grid. Therefore, fine-grained and advanced security protocols still need to be developed for protecting different communication networks in smart grids.

In a smart grid, the utility company needs the real-time power-consumption data for planning purposes as well as for providing accurate and authentic billing. For the utility company, the correctness of the calculated bills is the most important issue. However, from the customer's perspective, privacy is the main concern. Researchers have designed privacy-preserving billing protocols using advanced cryptographic techniques such as zero-knowledge proof (tinyurl.com/2z5blx) and homomorphic encryption (tinyurl.com/depohp). Bohil and colleagues (2010; tinyurl.com/mmfv5kt) proposed a privacy model for smart metering, in which a trusted third-party proxy is introduced to collect meter readings from individual customers and aggregate data before forwarding it to the utility company. Later on, Garcia and Jacobs (2012; tinyurl.com/kmxnnh7) proposed the use of homomorphic encryption to prevent the utility company from accessing the power consumption data of individual households. Using those advanced cryptographic techniques, utility companies only receive the commitments (tinyurl.com/ljgcasm) of the real-time power consumption instead of the raw data from smart meters, and customers can prove to the utility company that a utility bill has been correctly generated.

Besides research into addressing general privacy concerns for the smart grid, a number of researchers have been focusing on designing and implementing privacy-preserving billing protocols. Rial and colleagues (2011; tinyurl.com/lha6zpf) proposed a privacy-preserving billing protocol in which the power-consumption data is sent to the user along with other information from the smart meter, and the user computes the bill based on the pricing policy during each billing period. After that, the user sends the proof of correct computation to the utility company, where a homomorphic commitment scheme has been used to construct the proof. Kursawe and colleagues (2011; tinyurl.com/ldc7cfx) presented a set of protocols that can be used to privately compute aggregate meter measurements over defined sets of meters without revealing any additional information about the individual meter readings. Moreover, their protocols also allow for detection of fraud and leakage as well as network management and statistical processing of meter measurements. Molina-Markham and colleagues (2012; tinyurl.com/mjyz2a8) implemented the privacy-preserving billing protocol proposed by Rial on a MSP430-based microcontroller and verified the feasibility of designing privacy-preserving smart meters using low-cost microcontrollers.

Future Outlook

The smart-grid metering and control system consists of heterogeneous wired and wireless networks and devices from various domains. Each sub-system in the smart grid currently follows the different standards and regulations and has distinct security requirements. In particular, the smart grid faces unique challenges stemming from the combination of stringent security requirements, limited computational resources, time-critical message delivery and responses, and the use of heterogeneous networks with multiple authentication and protection mechanisms. Although a lot of efforts have been made by industry and academia to address a wide range of security issues in the smart grid, there are still many challenges that need to be tackled before smart grids can be widely deployed. From the viewpoint of cryptographic technique, we highlight several research areas and directions that need to be further investigated.

A lightweight cipher suite for smart-grid devices
The tight cost and resource constraints inherent in mass deployments of smart-grid devices bring forward impending requirements for implementing a lightweight cipher suite that can perform strong authentication and encryption, and provide other security functionalities. Previous research has shown that using classical cryptographic algorithms that are designed for full-fledged computers has become the bottleneck in many smart-grid applications. In order to meet the stringent time requirements in a smart grid, it is highly desirable to standardize a set of lightweight symmetric-key and asymmetric-key ciphers for securing smart-grid applications.

Advanced key management for smart-grid networks
Encryption and authentication are crucial cryptographic processes in a smart grid, because they protect data integrity and confidentiality, and an efficient key-management scheme is the foundation that ensures the secure operation of a smart grid. Because a smart grid is composed of heterogeneous communication networks and involves symmetric-key and asymmetric-key cryptosystems, a large set of cryptographic keys need to be managed in an efficient manner. A sophisticated key-management framework needs to be designed to deal with security services as well as the seamless handover of those services across different sub-systems in the smart grid.

Privacy-preserving operations in smart-grid networks
Smart-grid communications have raised serious concerns about user privacy due to the possibility of inferring customers' behaviour and habits from the detailed energy usage information, which can lead to potential risks that consumers would be vulnerable to criminal activities and personal information leakage. Advanced privacy-preserving security schemes need to be developed and integrated into smart-grid networks to enable utility companies to perform the regular business operations such as customer billing only using aggregated power-consumption information. The real-time power consumption data should only be accessible by individual customers.

Conclusion

Smart-grid metering and control systems hold enormous promise for improving efficiency, convenience, and sustainability. However, the complicated and heterogeneous system architecture has made securing the smart grid particularly challenging. Cybersecurity in the smart-grid metering and control system is an important and rapidly evolving area that has attracted attention from government, industry, and academia. In this article, we introduced the high-level architecture of a smart-grid metering and control system, detailed the system's security requirements, summarized the recent efforts from industry and academia, and highlighted several areas and directions for further research. Our objective is to shed some light on cybersecurity in the smart grid and to trigger the close collaborations among government, industry, and academia.

Based on our discussion in this article, it is clear that implementing an integrated and fine-grained security solution that is able to address potential security and privacy issues in each sub-system of a smart grid is critical to guarantee its successful deployment. Moreover, the design of security solutions should take into account the salient features of the smart grid as well as the underlying power system. Looking to the future, the joint efforts from industry and academia will make the era of “smart energy” become reality at a staggering speed.

Security Challenges in Smart-Grid Metering and Control Systems

Xinxin Fan and Guang Gong

About the Authors

Xinxin Fan is a Research Associate in the Department of Electrical and Computer Engineering at the University of Waterloo, Canada. He holds a PhD degree in Electrical and Computer Engineering from the University of Waterloo, as well as a BSc degree in Applied Mathematics and an MEng degree in Information Systems and Telecommunication Engineering from Xidian University, China. His research interests range from fast and secure software and hardware implementations of cryptographic algorithms to the design and the analysis of security protocols for wireless and wireline networks.

Guang Gong is a Professor in the Department of Electrical and Computer Engineering at the University of Waterloo, Canada, and she is the Managing Director of the Centre for Applied Cryptographic Research at University of Waterloo. She holds a BSc degree in Mathematics, an MSc degree in Applied Mathematics, and a PhD degree in Electrical Engineering from universities in China. Dr. Gong has also held a fellowship at the Fondazione Ugo Bordoni, in Rome, Italy, and was Associate Professor at the University of Electrical Science and Technology of China. Her research interests are in the areas of sequence design, cryptography, and communication security.

Citation: Fan, X. and G. Gong. 2013. Security Challenges in Smart-Grid Metering and Control Systems. *Technology Innovation Management Review.* July 2013: 42–49.

Keywords: smart grid, cybersecurity, privacy, encryption, authentication

Q&A

Sherif Koussa

Q. *Should startups care about application security?*

A. Many of the startup executives I meet think that application security is only for large companies, such as banks and government agencies. After all, these organizations have a lot of data to secure, established reputations to worry about, and trusted brands to protect. Startups do not have those things (yet), nor do they have the money to invest in anything that will not help them reach the next round of financing. They are focused on acquiring customers to establish better brand recognition. So, it is easy to understand why startups may be less than enthusiastic about the topic of application security. However, in my experience, many smart and successful CEOs and CTOs are not so quick to dismiss the topic. Startups that do not pay enough attention to security in the early stages may fail to later capitalize on the value of what they are building now. Furthermore, successful startup executives recognize the value of security as a market differentiator.

Unfortunately, it is not enough for startups to recognize that they need to care about application security; they need to take action. The challenge is cutting through the apparent complexity and building-in application security from the very beginning, while minimizing costs. Here, I will provide an overview of the key elements of application security, and I will discuss practical strategies that startups can use to increase the security of their applications throughout their lifecycles. I will focus on how a startup team can protect its application code against malicious threats, although a startup should also consider how security can provide opportunities to differentiate its application from the competition create opportunities in the marketplace.

Security Design and Architecture

When designing a secure application, security design and architecture is the key. There are six security cornerstones that must be kept in mind in every stage of the design:

1. **Confidentiality:** limiting access to data to only those who should have access to that data.

2. **Integrity:** ensuring that data has not been modified either accidentally or maliciously, either in transit or at rest.

3. **Availability:** ensuring that the data and the systems serving this data are up and running when needed.

4. **Authentication:** confirming the identity of a user or a system and proving that they actually are who they claim to be.

5. **Authorization:** ensuring that the authenticated entity has access rights to the resources that they claim they have access rights to.

6. **Non-repudiation:** proving whether or not an entity actually made a transaction they claim to have made.

Once these cornerstones have been established, there are three design concepts that come to play: attack resilience, attack tolerance, and attack resistance.

1. **Resistance:** the ability of the software to resist attacks. Principles that help with attack resistance include:

- *Defence in depth* (tinyurl.com/m5lkblc): building the security of a system in layers such that result is greater than the sum of its individual parts.

- *Attack surface* (tinyurl.com/5py7w4): minimizing the attack surface, which is those places where an attacker can start poking the application looking for holes.

- *Least privilege* (tinyurl.com/29a93a): giving users and processes only the minimum set of privileges to perform their function.

2. **Tolerance:** the ability of the software to tolerate failures. A principle that helps with attack tolerance is *failing securely* (tinyurl.com/h7vhm): a very important design principle that entails anticipating and handling exceptions in the software, so that the software does not end up in an insecure state in a fail scenario.

3. **Resilience:** the ability of the software to isolate attacks and contain the damage resulting from these attacks. A principle that helps with attack resilience is *compartmentalization*: an object-oriented programming concept that entails segregating different modules of the software. If a module is breached, it may be contained within that module and not necessarily spread to the whole application.

Increasing Awareness and Knowledge

The general strategies described above can help a startup improve its security posture, but they may still seem difficult to implement. However, it does not have to happen all at once. Software is developed in phases, and software security is built in the same way. The key is to take small yet measurable and progressive steps towards the goal. In many cases, the first step is for the startup to increase its staff's awareness and knowledge of security issues.

Companies should review their application-security awareness and security design. Even just knowing that an issue exists or is important can help a startup manage the associated risk. Ira Winkler (2012; tinyurl.com/acuofmc) argues that security awareness can be the most cost-effective security measure. Many code flaws happen because developers lack knowledge about proper secure coding and the reasons and consequences of writing a certain line of code in a certain way.

A great place to start increasing awareness and knowledge is by taking courses, either in person or online. I have also seen companies do very well with "lunch and learns" or similar in-house seminars with experts. Startup teams can also review lists of common security flaws, such as the "Top 10 list" published by the Open Web Application Security Project (tinyurl.com/3n6q9rg), both to increase awareness and assess their own application's security. Deliberately insecure applications in various languages are also available for testing and learning purposes; an example is WebGoat (tinyurl.com/62kggay) for Java-based web applications.

Taking Action Through Controls

Once security awareness has been established, safeguards or countermeasures must be put in place to ensure that the knowledge obtained during the awareness phase is actually implemented in the code. Usually, application-security controls are divided between preventative and detective controls:

1. **Preventative controls:** These controls include the security awareness implemented in the previous phase and other controls. Examples include:

- *Security checklists*: Checklists are the most effective security controls, yet their value is frequently underestimated and they are underused. A security checklist is simply a list of all the things a developer should check before committing code to the repository. Helpful resources include Mozilla's Secure Coding Guidelines (tinyurl.com/4ynfbqn) and Secure Coding QA Checklist (tinyurl.com/km3et2m), as well as MSDN's Secure Coding Guidelines (tinyurl.com/67a6ne9). Also, Patch++ (patchplusplus.com) provides a visual way to implement checklists for securing code patches.

- *Security code review*: There are many flavours of security code review, but the simplest form is a regular peer code review infused with security guidelines and checks developed from the security checklists mentioned above. The most inclusive form is a full-scale security code review involving the use of automated tools and scripts as well as manual inspection of the code. Security code review is one of the best controls for a software development lifecycle; it can prevent the largest number of security flaws from making it to production, and it provide the quickest means of remediation. For my simplified version of a security code review process, see Koussa (2013; tinyurl.com/kbhwy3s).

2. **Detective Controls**: The two most-common detective controls in application security are:

- *Penetration testing*: with this control, an internal or external security analyst tries to emulate what an attacker would do to look for vulnerabilities in a given piece of software and then tries to exploit them. Other names for this type of control include vulnerability assessment, vulnerability scanning, or dynamic testing, but they all represent more or less the same type of control with different levels of thoroughness. Penetration testing is a very good control to measure the "hackability" of the application.

- *Web-application firewalls*: these are firewalls that monitor traffic going in and out of a web application. Depending on the firewall's capabilities, the firewall could potentially block inputs and outputs that do not meet the criteria defined in its set of rules. Web-application firewalls are often a popular option to protect deprecated or soon-to-be-deprecated applications. They are also a popular choice to provide some protection for applications that are deemed too costly to fix.

Q&A. Should Startups Care about Application Security?

Sherif Koussa

Implementing Processes

Once a startup team is aware of the security threats, risks, and attacks that are relevant to their application, once they knows what needs to be done in order to counter these attacks, and once they are implementing a few controls to ensure that the awareness is practically implemented, the next phase is to implement a systematic and measurable process across all disciplines of the software development lifecycle to ensure that fewer and fewer vulnerabilities make it to production. There are several approaches to securing software-development lifecycles, such as Microsoft SDL (tinyurl.com/y6frgge), or initiatives that help integrate security into existing models, such as BSIMM (bsimm.com) and OpenSAMM (opensamm.org).

The challenge for any process is whether it actually is adopted by the development teams, who may not welcome adding additional processes if they perceive process to interfere with the actual job of writing code (Turner, 2011; tinyurl.com/44xh5sw). When it comes to choosing a secure software development lifecycle process or introducing new security activities into existing ones, I always suggest small yet progressive steps. Nothing is more damaging than to shock development teams by suddenly imposing heavy processes.

Conclusion

Startups can no longer afford to ignore application security. It is not a question of whether or not startups should care about application security; they need to do more than care – they need to take action. However, taking effective steps toward secure software does not have to come with a hefty drain on the startup’s budget or productivity levels. On the contrary, some startups are using software security as a marketing differentiator in an age when clients are looking for more privacy and demanding evidence of privacy controls implemented by the organization.

Recommended Reading

- "Application Security Architecture" (Simhadri, 2001; tinyurl.com/nyu7lzc)
- Software Security Engineering: A Guide for Project Managers (Allen et al., 2008; tinyurl.com/lua92tb)
- "Architecture and Design Considerations for Secure Software" (SwA Forum and Working Groups, 2012; tinyurl.com/mmx928h)

About the Author

Sherif Koussa is Principal Application Security Consultant and founder of Software Secured, an application security firm. He has spent 14 years in the software development industry, with the last six years focused on testing application security, assessing security, and teaching developers to write secure code. He worked on the OWASP security teaching tool WebGoat 5.0, helped SANS launch their GSSP-JAVA and GSSP-NET programs, and wrote the blueprints of the Dev-544 and Dev-541 courses. In addition, he authored courseware for SANS SEC-540: VOIP Security. Sherif leads both the OWASP Ottawa Chapter and the Static Analysis Code Evaluation Criteria for WASC. He has performed security code reviews for three of the five largest banks in the United States. Before starting Software Secured, Sherif worked on architecting, designing, implementing, and leading large-scale software projects for Fortune 500 companies, including United Technologies, and other leading organizations such as Nortel Networks, March Healthcare, Carrier, Otis Elevators, and NEC Unified Communications.

Citation: Koussa, S. 2013. Q&A. Should Startups Care about Application Security? *Technology Innovation Management Review.* July 2013: 50–52.

Keywords: cybersecurity, application security, software security, training, checklists, startups, code reviews, detection, prevention, design, architecture

TIM Lecture Series

Green Business Models to Change the World: How Can Entrepreneurs Ride the Sustainability Wave?

Mika Westerlund

> *"We need entrepreneurs and leaders with the courage and conviction to take bold action ahead of others. We also need radically new business models that create true value for the environment and society, bring competitive advantage to companies, and have the potential to transform industries globally."*
>
> Mika Westerlund
> Assistant Professor
> Sprott School of Business

Overview

The fifth TIM lecture of 2013 was presented by Mika Westerlund, Assistant Professor at Carleton University's Sprott School of Business (sprott.carleton.co) in Ottawa, Canada. Westerlund discussed the need for a new sustainability-oriented business culture; described emerging business models that aspiring entrepreneurs can create or adopt; and presented recent research and trends relating to sustainability and green innovation. The event was held at Carleton University on June 20th, 2013.

The TIM Lecture Series is hosted by the Technology Innovation Management program (carleton.ca/tim) at Carleton University. The lectures provide a forum to promote the transfer of knowledge from university research to technology company executives and entrepreneurs as well as research and development personnel. Readers are encouraged to share related insights or provide feedback on the presentation or the TIM Lecture Series, including recommendations of future speakers.

Summary

In the first part of the lecture, Westerlund focused on the need to shift mindsets toward sustainability and the ways in which this can be accomplished. In the second part, he focused on the mechanisms by which companies can profit from a focus on sustainability.

Part I: Value creation

When viewed through a traditional mindset, the goals of "green" and the goals of "business" seem incompatible; however, companies that embrace a green mindset are becoming increasingly successful in today's economy. For entrepreneurs who have embraced a sustainability mindset, the compatibility of *profit* and *planet* may be combined with an emphasis on *people* and *personal* benefits; these are the "four Ps" of sustainable entrepreneurship (GEF, 2011; tinyurl.com/mldfyqe). Thus, entrepreneurs are passionate about making a positive impact on their environment, their society, and their economy.

Sustainable entrepreneurship is not primarily about starting a business but about taking responsibility for life choices and promoting this way of thinking. However, this does not mean that the focus on sustainability must come at the expense of profit. If fact, this new mindset can open an entrepreneur's eyes to opportunities that others may not see, whether it is an idea for a new venture, an opening in the market, or a more sustainable process that can be applied to an existing business.

Green Business Models: How Can Entrepreneurs Ride the Sustainability Wave?

Mika Westerlund

To varying degrees, green companies are taking advantage of five key benefits of a focus on sustainability (Kiron et al., 2013; tinyurl.com/pdgbtnr):

1. **Market benefits** (e.g., brand, competition, new markets)

2. **Financial benefits** (e.g., increased margins, reduced costs)

3. **Innovation benefits** (e.g., business models and processes, product/service offerings)

4. **Compliance benefits** (e.g., reduced waste, lower material costs, adherence to regulations)

5. **Stakeholder benefits** (e.g., attracting and retaining talent, stakeholder/investor relations, reduced risk)

Among the market benefits, a key driver has been the increasing demand and prices of natural resources, therefore the greatest impetus for green innovation lies in resource-intensive industries that produce consumer products, chemicals, automobiles. Here, green innovation is not just about "doing good"; a company in a resource-intensive industry simply cannot compete without going green (Haanaes et al., 2012; tinyurl.com/acawzkp). Indeed, the green mindset can be applied to any aspect of a company, including production processes, lifecycle management, new products and services, and new business models.

The more intangible market benefits are also important, including the brand benefits. Positive perceptions of a brand based on its eco-credentials can be of great benefit to a company. However, despite the fact that sustainability brands can be relatively easy to build, they are also easily damaged, whether through the company's own actions or the actions of others (e.g., competitors, activists, disgruntled customers). To avoid accusations of "greenwashing" (tinyurl.com/nlvbfrg) and to demonstrate its commitment to the principles of sustainability, transparency is important.

Despite the various recognized benefits, the need for sustainability is not recognized worldwide. North America, in particular, lags well behind other economies in terms of business-model innovation and investment in sustainability; emerging markets such as Africa, the Middle East, and Asia-Pacific region, lead the way (Kiron et al., 2013; tinyurl.com/pdgbtnr).

However, even in North America, incremental improvements are making an impact. To illustrate the benefits of incremental green innovation, Westerlund provided examples of eco-efficiency in the following domains:

1. **Data centres:** increased energy efficiency and investment in renewable energy technologies (e.g., Facebook, Google, Intel)

2. **HVAC systems:** optimized electrical usage in commercial buildings (e.g., LOBOS system; enerliance.com/lobos/)

3. **Solar-powered airplanes:** reduced fuel usage or even fuel-free flights (e.g., Solar Impulse; solarimpulse.com)

4. **Airlines:** fuel-efficiency targets and innovations (e.g., Virgin Atlantic; tinyurl.com/ohj6329)

5. **Beverage companies:** improved manufacturing processes (e.g., Coca-Cola and World Wildlife Fund; tinyurl.com/oro2apx)

6. **Other examples**: smart renewable technology, renewable energy, clean tech, smart grids

Although eco-efficiency is beneficial, related innovations are incremental in nature, are easily copied, and do not on their own enable a company to become a green champion. For companies to gain a distinct competitive advantage, they must change the established ways of thinking, disrupt the market, and transform the industry practices and business models. Thus, Westerlund also provided examples of radical, game-changing innovations. The focus was on Interface (tinyurl.com/d92kjd), a global carpet manufacturer that has pursued a bold and financially successful vision for sustainability. The Interface story was recently featured in the TIM Review through a case study by Lampikoski (2012; timreview.ca/article/624).

Next, Westerlund summarized what can be learned from those who have embraced sustainability and have pursued green innovations:

1. **Move early, even if information is incomplete:** sustainability is an evolutionary process with multiple stages. The journey should be initiated as a reaction to growing risks and uncertainties, and it is characterized by discoveries.

2. **Balance the short- and long-term benefits:** set a broad, long-term vision with projects offering concrete, near-term “wins”.

3. **Drive sustainability from top-down and bottom-up:** enlist employees at multiple levels for improved results and engagement, listen to staff who are aware of sustainability, gather ideas, promote cultural change, make staff feel proud.

4. **Aggressively de-silo sustainability:** the approach should be integrated throughout company operations; build sustainability into core processes (and partners’ processes).

5. **Measure everything:** if ways of measuring something do not exist, start inventing them.

6. **Value intangible benefits:** a meaningful portion of a sustainability strategy may relate to intangible benefits.

7. **Be authentic and transparent:** be realistic, communicate challenges and success, stress long-term goals over short-term goals.

Westerlund also emphasized that small firms are better able to embrace sustainability than large firms, for the following reasons:

1. Small firms innovate; large firms bring innovation to masses. Startups build; incumbents transform.

2. Networks favour small firms, and radical innovation is associated with startups.

3. Due to pressure from investors, large firms are often limited to ensuring success through incremental innovation.

4. Small firms are flexible in implementing business models that break the industry rules. Creative destruction is easier for startups, because they are more agile and less encumbered by change management.

5. Large firms need more time to adopt change in strategy.

Overall, the main messages from the first part of the lecture were: a) that value creation depends on cultural change and b) that change is in the air. Increasingly, companies are coming around to a sustainable way of thinking, and they are benefiting from this change in mindset. However, this change does not happen overnight, whether within a specific company or industry.

Part II: Value capture

In the second part the lecture, Westerlund focused on sustainable business models, which may be entirely new or they may be simply modifications of existing business models. Generally, it can be said that, the more parts of a business model which are changed and have a green effect, and the more profoundly a green change is taking place within the individual parts of the business model, the greener the business model innovation and the higher potential for creating radical eco-innovation (Henriksen et al., 2012; tinyurl.com/oagsn65).

Westerlund provided descriptions and examples of the following types of sustainable business models:

1. **Cause-related models:** tie the business model to a particular cause – such as saving the planet, curing a disease, or providing shoes, prescription glasses, or related medical treatment to regions in need – that will resonate with customers (e.g., Patagonia: patagonia.com; (RED): joinred.com; TOMS: toms.com).

2. **Functional sales:** provide "product-as-a-service" (e.g., rental/leased offerings, recycling of old products).

3. **Waste management:** reduce waste and lower costs for customers by providing management and supply contracts (e.g., chemical management/procurement systems).

4. **Energy services:** optimize energy usage for customers and be paid according to performance/savings (e.g., energy management for public buildings or industrial companies, software companies providing solutions to support the energy efficiency of their customers, residential solar systems).

5. **Sharing:** provide access to products, tools, shelter, and other resources rather than selling them as products (e.g., tool libraries, bike/car sharing, coworking office space)

6. **Re-using and recycling:** turn waste products into new products (e.g., recycled clothing, fashion products, hardware).

7. **Design on demand:** seek out inefficient products where "on demand" designs would be profitable (e.g., 3D garment printers).

8. **Hybrid models:** many sustainable business models are combinations of the models described above.

To conclude the lecture, Westerlund offered his view of the next steps for sustainability:

1. Stronger focus on sustainable business models: from types to design to management, and including the capabilities needed to manage the change

2. Focus on sustainability ecosystems: sustainable value networks fighting against each other, and the capabilities needed to manage ecosystems

3. Open and user innovation for green innovations: for example, living labs for SmartCities, energy efficiency, bottom-of-the-pyramid solutions

Lessons learned
In the discussions that followed each portion of the presentation, audience members shared the lessons they learned from the presentation and injected their own knowledge and experience into the conversation.

The audience identified the following key takeaways from the presentation:

1. Sustainability requires vision and systematic change. Developing the right mindset is a key success factor.

2. Compared with other parts of the world – particularly developing countries – North American companies are falling behind in sustainability; they are not investing enough money in sustainability nor are they giving it sufficient attention. From a company perspective, there does not appear to be a sense of urgency in North America.

3. Incremental innovation results in small or minor improvements. Companies need to do more to provide good, sustainable contributions. Radical innovation results in a game-changing innovation that involves both technology innovation and business model innovation.

4. Eliminating waste can lead to increasing wealth.

5. A sustainability brand is vulnerable to perceptions, and damage to one of the company's sustainability brands can cause further damage to the company overall.

6. Intellectual property protection may not be relevant for many sustainability innovations. The need for transparency and collaboration with competitors may be stronger than the need for patent protection.

7. Rethink your partnerships and your networks; have an "ecosystem view" of sustainability.

8. The competition is not company against company; rather, it is ecosystem against ecosystem.

9. There are many opportunities for entrepreneurs and small companies. Small companies are more agile than large companies, they can move faster, and they can fast-track sustainability. In this way, sustainability may represent a risk to larger companies.

10. Entrepreneurs and companies should focus on sustainable business models, sustainable ecosystems, and open/user innovation to create green innovations.

11. A good first step for companies is to identify and then replicate an existing (and successful) sustainability business model that is applicable to their type of business.

12. There are no measures or indicators (generally accepted or standards) for sustainability.

13. To overcome misuse of terms and bogus claims on sustainability, government policies and standards are required. You must be able to prove you are green.

14. Sustainability must become part of your whole business. For startups, this process must start on day one.

15. We need to rethink business, intellectual property systems, and licensing of green intellectual property. We must become stewards of sustainability.

Green Business Models: How Can Entrepreneurs Ride the Sustainability Wave?

Mika Westerlund

This report was written by Chris McPhee; the lessons learned were captured by Derek Smith.

About the Speaker

Mika Westerlund, D.Sc. (Econ.) is an Assistant Professor at Carleton University's Sprott School of Business in Ottawa, Canada. He previously held positions as a Postdoctoral Scholar in the Haas School of Business at the University of California Berkeley and in the School of Economics at Aalto University. Mika earned his doctoral degree in Marketing from the Helsinki School of Economics. His doctoral research focused on software firms' business models and his current research interests include open innovation, business strategy, and management models in high-tech and service-intensive industries.

Citation: Westerlund, M. 2013. TIM Lecture Series – Green Business Models to Change the World: How Can Entrepreneurs Ride the Sustainability Wave? *Technology Innovation Management Review.* July 2013: 53–57.

Keywords: sustainability, green innovation, business models, value creation, entrepreneurship, eco-efficiency

Author Guidelines

These guidelines should assist in the process of translating your expertise into a focused article that adds to the knowledge resources available through the *Technology Innovation Management Review*. Prior to writing an article, we recommend that you contact the Editor to discuss your article topic, the author guidelines, upcoming editorial themes, and the submission process: timreview.ca/contact

Topic

Start by asking yourself:

- Does my research or experience provide any new insights or perspectives?
- Do I often find myself having to explain this topic when I meet people as they are unaware of its relevance?
- Do I believe that I could have saved myself time, money, and frustration if someone had explained to me the issues surrounding this topic?
- Am I constantly correcting misconceptions regarding this topic?
- Am I considered to be an expert in this field? For example, do I present my research or experience at conferences?

If your answer is "yes" to any of these questions, your topic is likely of interest to readers of the TIM Review.

When writing your article, keep the following points in mind:

- Emphasize the practical application of your insights or research.
- Thoroughly examine the topic; don't leave the reader wishing for more.
- Know your central theme and stick to it.
- Demonstrate your depth of understanding for the topic, and that you have considered its benefits, possible outcomes, and applicability.
- Write in a formal, analytical style. Third-person voice is recommended; first-person voice may also be acceptable depending on the perspective of your article.

Format

1. Use an article template: .doc .odt

2. Indicate if your submission has been previously published elsewhere. This is to ensure that we don’t infringe upon another publisher's copyright policy.

3. Do not send articles shorter than 1500 words or longer than 3000 words.

4. Begin with a thought-provoking quotation that matches the spirit of the article. Research the source of your quotation in order to provide proper attribution.

5. Include a 2-3 paragraph abstract that provides the key messages you will be presenting in the article.

6. Only the essential references should be included. The URL to an online reference is preferred; where no online reference exists, include the name of the person and the full title of the article or book containing the referenced text. If the reference is from a personal communication, ensure that you have permission to use the quote and include a comment to that effect.

7. Provide a 2-3 paragraph conclusion that summarizes the article's main points and leaves the reader with the most important messages.

8. Include a 75-150 word biography.

9. If there are any additional texts that would be of interest to readers, include their full title and location URL.

10. Include 5 keywords for the article's metadata to assist search engines in finding your article.

11. Include any figures at the appropriate locations in the article, but also send separate graphic files at maximum resolution available for each figure.

Issue Sponsor