Technology Innovation Management Review

April 2017
Volume 7 Issue 4

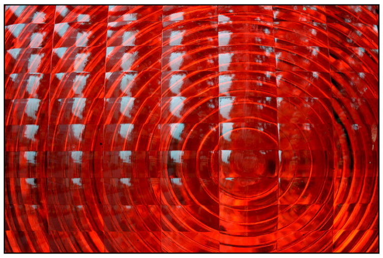


Image credit: XoMEoX (CC-BY)

Cybersecurity

Welcome to the April issue of the *Technology Innovation Management Review*. 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

Technology Innovation Management Review

April 2017
Volume 7 Issue 4

Publisher

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

ISSN

1927-0321

Editor-in-Chief

Chris McPhee

www.timreview.ca

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

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 articles.
- 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

About TIM

The TIM Review has international contributors and readers, and it is published in association with the Technology Innovation Management program (TIM; timprogram.ca), an international graduate program at Carleton University in Ottawa, Canada.

Editorial: Cybersecurity

Chris McPhee, Editor-in-Chief

Michael Weiss, Guest Editor

From the Editor-in-Chief

Welcome to the April 2017 issue of the *Technology Innovation Management Review*. This month's editorial theme is **Cybersecurity**, and I am pleased to welcome **Michael Weiss**, Associate Professor in the Department of Systems and Computer Engineering and the Technology Innovation Management (TIM; timprogram.ca) program at Carleton University in Ottawa, Canada.

In May, we examine the theme of **Lean and Global** with Guest Editor **Stoyan Tanev**, Associate Professor of Innovation & Design Engineering at the University of Southern Denmark.

For future issues, we are accepting general submissions of articles on technology entrepreneurship, innovation management, and other topics relevant to launching and growing technology companies and solving practical problems in emerging domains. Please contact us (timreview.ca/contact) with potential article topics and submissions.

Chris McPhee
Editor-in-Chief

From the Guest Editor

The articles in this special issue mirror some of the recent developments in cybersecurity. The Internet of Things and Internet-enabled medical devices are changing the security landscape: i) cyber attacks can be carried out on a much larger scale by levering devices that have less computing power and are, therefore, harder to protect against cyber-attacks, and ii) attacks can also affect humans lives directly through medical devices that are accessible via the Internet and embedded into the human body. A third area explored by the articles is the connection between cyber security and big data.

In the first article, **Mikko Hypponen**, Chief Research Officer at F-Secure, and **Linus Nyman**, Assistant Professor at the Hanken School of Economics in Helsinki, Finland, highlight the importance of security engineering for manufacturers building devices for the Internet of Things (IoT). Building on Hypponen's law, which asserts that "Whenever an appliance is described as being 'smart', it's vulnerable.", the authors offer recommendations to help manufacturers and consumers address the vulnerabilities of smart devices. They also highlight the importance of legislation in securing the Internet and its connected devices.

Next, **Mackenzie Adams**, Co-Founder and Creative Director at SOMANDA Inc., examines individual privacy in the IoT, specifically as it relates to big data. Drawing on evidence from recent big data breaches, the authors assert that the collection of data from IoT devices, and subsequent customization based on the collected data, create vulnerabilities in individual data privacy. The article examines the complexity of tackling technological and legislative challenges in protecting individual privacy. The authors position these issues in terms of the future implications of the IoT and the loss of privacy.

Then, **Ahmed Shah** and **Michael Weiss** from Carleton University; **Ibrahim Abualhaol** from Larus Technologies; and **Mahmoud Gad** from the VENUS Cybersecurity Corporation, describe the creation of a prototype system for monitoring real-time Border Gateway Protocol (BGP) traffic for security threats. By combining

Editorial: Cybersecurity

Chris McPhee and Michael Weiss

modes of exploratory analysis and automated analysis, the system enables security analysts to discover new anomalies and validate detection rules.

Finally, **Aida Alvarenga** and **George Tanev** from the Technology Innovation Management program at Carleton University propose a cybersecurity risk-assessment framework that integrates value-sensitive design. Using the field of medical devices as a case domain in which to ground their framework, the authors review the relevant literature through the perspective of using security initiatives as a value proposition that could be communicated to the medical device manufacturer’s stakeholders. To illustrate how it can be applied to a device and used to select the risk controls that bring the most value to the device’s key stakeholders, they apply their framework to the theoretical case of an insulin pump.

We hope that you enjoy reading this issue and will learn about some of the recent trends in cybersecurity.

Michael Weiss
Guest Editor

About the Editors

Chris McPhee is Editor-in-Chief of the *Technology Innovation Management Review*. He holds an MASc degree in Technology Innovation Management from Carleton University in Ottawa, Canada, and BScH and MSc degrees in Biology from Queen's University in Kingston, Canada. Chris has nearly 20 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.

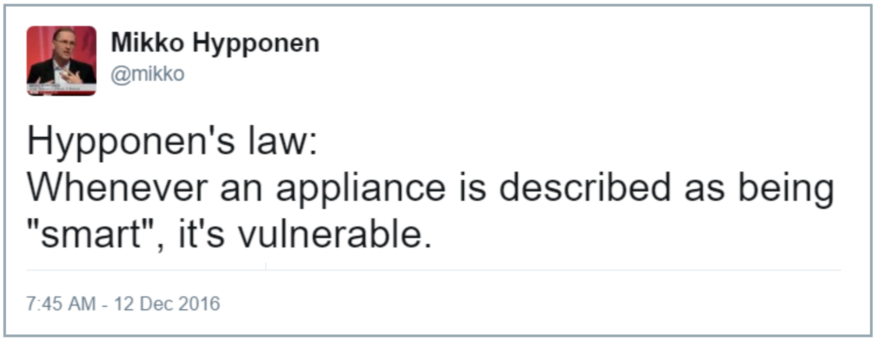
Michael Weiss holds a faculty appointment in the Department of Systems and Computer Engineering at Carleton University in Ottawa, Canada, and is a member of the Technology Innovation Management program. His research interests include open source, ecosystems, mashups, patterns, and social network analysis. Michael has published on the evolution of open source business, mashups, platforms, and technology entrepreneurship.

Citation: McPhee, C., & Weiss, M. 2017. Editorial: Cybersecurity. *Technology Innovation Management Review*, 7(4) 3–4.
http://timreview.ca/article/1065

Keywords: cybersecurity, Internet of Things, IoT, Hypponen’s law, vulnerabilities, smart devices, security engineering, privacy, big data, legislation, anomaly detection, real time, exploration, automation, risk assessment, medical devices, value proposition

The Internet of (Vulnerable) Things: On Hypponen's Law, Security Engineering, and IoT Legislation

Mikko Hypponen and Linus Nyman



Mikko Hypponen
@mikko

Hypponen's law:
Whenever an appliance is described as being "smart", it's vulnerable.

7:45 AM - 12 Dec 2016

The Internet of Things (IoT) and the resulting network-connectedness of everyday objects and appliances in our lives bring not only new features and possibilities, but also significant security concerns. These security concerns have resulted in vulnerabilities ranging from those limited in effect to a single device to vulnerabilities that have enabled IoT-based botnets to take over hundreds of thousands of devices to be used for illegal purposes. This article discusses the vulnerable nature of the IoT – as symbolized by Hypponen's law – and the parts both manufacturers and consumers play in these vulnerabilities. This article makes the case for the importance of security engineering for IoT manufacturers, highlights some significant issues to help consumers address these vulnerabilities, and argues for legislation as perhaps the only reliable means of securing the Internet and its connected devices.

Introduction

As security expert Bruce Schneier (2015) has noted, the appliances and gadgets that are part of our everyday lives are becoming computers that can do other things. Our phones have become computers that can also make phone calls. Our cars are becoming computers that can also drive. Our washing machines are becoming computers that can also wash clothes. These computers are commonly connected to a network – often, though not necessarily, the Internet. The phenomenon as a whole is called the Internet of Things (IoT; tinyurl.com/lqdsl4n). Between 2014 and 2020, the number of these connected things has been projected to grow at an annual compound rate of 23.1%, reaching 50.1 billion things in 2020 (Press, 2016).

This emerging ubiquity of network-enabled computers raises a host of significant privacy and security concerns. As Chief Research Officer for F-Secure, a Finnish cybersecurity company, this article's main author has spent more than a quarter of a century working to make computers safe. As the first IoT devices, or "smart" devices, began appearing on the market, Hypponen, along with many other security experts, began taking a closer look at them. The results were very worrying indeed: these connected devices almost invariably contained significant vulnerabilities.

The vulnerable nature of network-connected devices has been covered before in both the popular press (e.g., Franceschi-Biccierai, 2016a; Greenberg & Zetter, 2015; Schneier, 2014) as well as in academia (e.g., Abomhara & Køien, 2015; Greene, 2015; Patton et al., 2014). Particularly within the security community, these topics have been discussed and warned about for years. And yet, both in the popular press as well as among security researchers, there are many who believe the situation regarding IoT vulnerabilities is getting worse, not better (e.g., Franceschi-Biccierai, 2016b; Porup, 2016; Schneier, 2017). IoT vulnerabilities have been shown to affect not only the quality of individual products and networks, but also even the stability of the very backbone of the Internet itself. By extension, these vulnerabilities impact the wellbeing of human life as well.

This article is primarily for readers with limited to no experience in security engineering. It is part academic essay on the vulnerable nature of the Internet of Things and part plea to manufacturers and consumers to take these vulnerabilities seriously. We believe it will be of particular interest to three main groups. First, to managers or manufacturers who are considering entering the world of IoT. Second, to consumers who want to better understand some of the risks of their smart products and how to mitigate them. And, third, to legislators concerned with the safety and security of our everyday devices.

The remainder of this article is structured as follows. We begin with a brief discussion of the rise of the IoT and its part in transforming traditional companies into software companies. We then examine the vulnerable nature of smart devices, provide examples of vulnerabilities, and discuss some key reasons why these vulnerabilities exist. Finally, we recommend actions that can be taken by both manufacturers and consumers to address these vulnerabilities, and we conclude with a brief discussion of legislation as a means of securing the IoT.

IoT: Old Concepts, New Software Companies

The Internet of Things as a phenomenon is not new. In 2014, the IoT made the top of Gartner's list of the most hyped emerging technologies (Gartner, 2015). However, the concepts that form the building blocks of the IoT are considerably older; the phenomenon itself is made possible by half a century of advances in computing. Among the more significant changes over the past decade that have enabled IoT's meteoric rise are a significant drop in cost for the necessary component parts of smart devices and the widespread availability of Wi-Fi. In other words, getting things online is becoming very inexpensive and getting them connected is becoming very easy.

Although there are notable challenges to monetizing the IoT (e.g., Westerlund et al., 2014), predictions about the IoT being headed for massive growth are the norm. Over the longer term, some believe its growth will surpass even that of the early Internet (e.g., Gershenfeld & Vasseur, 2014). Over the shorter term, estimates for both IoT market size and growth are also substantial. McKinsey puts IoT market size estimates at increasing from $900M USD in 2015 to $3.7B in 2020 (e.g., Forbes, 2016), and Bain predicts that, by 2020, the annual revenue for vendors of IoT hardware, software, and "comprehensive solutions" may exceed $470B (Forbes, 2016).

Indeed, we have already seen companies take strong strategic stances in support of the IoT. Samsung Co-CEO Boo-Keun Yoon proclaimed, back in 2015, that 90% of Samsung products would be IoT-enabled by 2017, and 100% by the year 2020 (Sims, 2016). Yoon did not state that Samsung would add IoT-enabling components to all those products where an Internet connection would offer some consumer benefit. Rather, that it would make *all* of its products IoT-enabled. And Samsung is not alone. Even a brief glance at the plethora of smart products flooding the market suggests that there is an abundance of companies striving to make anything and everything IoT-enabled. The resulting spectrum of IoT devices covers everything from more self-evidently useful implementations such as smart security cameras to increasingly odd, even bizarre, implementations including toasters (Vanhemert, 2014), mattresses (Crook, 2016), showerheads (Krupitzer, 2015), and underwear (Graham, 2016).

Consumers may not see the benefits of an Internet connection in all of their devices. IoT features may, instead, be intended to benefit the company that produces them, in the form of collected data. Data was, of course, considered a crucial topic even before the emergence of IoT. (In fact, when IoT made Gartner's [2015] list of the most hyped technologies, it did so by displacing "Big Data".) IoT devices are in a unique position to gather data for their manufacturers about the product's use: how often we wash our clothes, how many cups of coffee we drink each day, and so forth. In an effort to, in part, offer products with new IoT features, but also in an effort to gather additional valuable data, numerous companies that just a few years ago had nothing to do with software are now rushing to join the IoT revolution – and, in the process, are becoming software companies. A significant reason why this shift is problematic, and indeed the underlying cause behind so many of the vulnerabilities we see today, is the resulting lack of experience in security engineering among these new software companies.

The Internet of (Vulnerable) Things

Mikko Hypponen and Linus Nyman

Hypponen's Law: Smart Means Vulnerable

Hypponen's law is a simple yet important concept – so simple, in fact, that it was first put forth as a single tweet in December 2016 (http://twitter.com/mikko/status/808291670072717312) "Hypponen's law: Whenever an appliance is described as being 'smart', it's vulnerable." Whether it is a car, a TV, or a toothbrush, if it is smart – if it is connected to a network – then it is vulnerable. This notion of the vulnerability of smart objects is of course not limited to appliances, but is equally true of other Internet-enabled things. Indeed, the ever-growing list of IoT devices ranges from mousetraps (Corfield, 2017) and tea kettles (Bode, 2015) to sniper rifles (Greenberg, 2015), cars (Greenberg, 2016), and beyond.

Our hope with this article is to reach out beyond the confines of the security community to further underline the simple yet important point of IoT vulnerability. If you are in the market for a smart product, you will be buying a vulnerable product. If you are designing a smart product, you are designing a vulnerable product.

The Far-Reaching Effects of IoT Vulnerabilities

Vulnerabilities can have very real and very bad results. A vulnerable IoT device can become a bridge between a private network and a public one. A vulnerable IoT device can be exploited to gain sensitive information, including passwords. The network-connectedness of IoT devices can serve as a means for malware to access not only the IoT device itself, but also other devices connected to the network. IoT vulnerabilities can even have consequences that extend far beyond the scope of a single device or local area network. This was the case in October of 2016, when large parts of the backbone of the Internet came under the largest attack in the history of the Internet. This attack was not conducted by supercomputers, or indeed even powerful desktop computers – it was conducted by over 100,000 IoT appliances. These appliances, unbeknownst to their owners, became part of the "Mirai botnet", whose initial targets ranged from an individual security journalist (Krebs, 2016a) to several waves of attacks against a company that provides core Internet services for dozens of popular sites, among them Twitter, Spotify, Reddit, and the New York Times (Etherington & Conger, 2016; Krebs, 2016b; Newman, 2016). This latter attack brought down a significant portion of the Internet for several hours.

Connecting things to the Internet can lead to vulnerabilities for reasons unrelated to the devices themselves. An example of this is an industrial control system interface that has been connected to the Internet without including security measures such as requiring the user to log in or enter a password. These kinds of interfaces may have been connected to the Internet intentionally but then security measures, such as requiring a password, were forgotten to be implemented. Alternatively, an interface may have initially been set up on a separate network that was not connected to the Internet. Then, perhaps several years later, that network was connected to the Internet, without those who connected it having realized that connecting it made the industrial control system interface accessible to anyone on the Internet. For example, security researchers at F-Secure have discovered such unsecured systems that control prescription drug orders, home automation and security systems (to control temperature, security cameras, alarms, and even curtains), car washes, pumping stations, swimming pools, restaurant point-of-sale systems, solar panels, biogas plants, ski lifts, wind turbines, hospital bed monitoring stations, funeral parlour crematoriums, and steel furnaces.

Why Is Smart Vulnerable?

There are two basic causes of IoT vulnerabilities: technical problems and people problems. In the following subsections, we discuss each type of problem individually.

Technical problems

By technical problems, we mean problems that can be fixed with an update. There will never come a time when new vulnerabilities are no longer discovered, and therefore the security of any system depends on that system being kept up-to-date. People are notoriously poor at regularly updating their systems – this is a "people problem" – which is why automatic updates have become common. One significant problem in IoT devices is that it may be difficult, or even impossible, to update their software. Both the operating system and the software running the IoT device must be updateable. If they are not, or even if updating one or both of them is not easy, the emergence of exploitable vulnerabilities in a product is a near certainty. To make matters worse, some IoT devices ship with outdated operating systems, meaning the devices may have known vulnerabilities before they are even unboxed.

In addition to outdated software, a further significant source of vulnerabilities stems from the failure of IoT manufacturers to take advantage of lessons already learned by others. Many technical problems have been solved years ago, even decades ago, resulting in evolving sets of best practices in the computer industry. However, vulnerabilities that should no longer be a problem continue to plague the IoT. An example of this is the Telnet communications protocol. Telnet is an unsecured means of communicating over a network. Due to its lack of encryption, the computer industry moved away from Telnet roughly two decades ago. However, Telnet can still be found among the causes of current IoT vulnerabilities (e.g., Franceschi-Bicchierai, 2016c; Krebs, 2016).

People problems
Whereas technical problems typically can be fixed with an update, people problems require education and learning, as well as an interest in addressing the issue or problem. In theory, people problems should be the easier of the two to fix. In practice, however, this is rarely the case. For example, consider the VHS recorder clock display. Readers old enough to remember the VCR are likely to have come across displays that showed a blinking “12:00” rather than the current time. In the case of the VCR, the effects of the user not making an effort to learn how to set the time were insignificant. However, this same phenomenon of user ignorance or indifference in the context of IoT appliances has a much greater impact. A key example of this is device default passwords. The Mirai botnet, for instance, was designed to search the Internet for IoT devices, trying a number of different common default usernames and passwords in order to gain control of the devices it found (e.g., Franceschi-Biccierai, 2016c). Something as simple as changing the default password on a device would have protected against this attack. We as users need to both know that default passwords are insecure and then also care enough about the issue to change them. A device capability, including security capabilities such as the ability to change a password, can be made ineffective through user ignorance or indifference.

Towards a More Secure IoT

It is our sincere hope that, ten years from now, we will be able to say about the IoT revolution what we can now say about the Internet revolution: the good outweighed the bad. However, this result will not come about by itself – concrete action is needed to curb IoT vulnerabilities. In the remainder of this article, we discuss some steps manufacturers, consumers, and legislators can take to mitigate IoT vulnerabilities.

Manufacturers
It is not our goal with this article to offer a checklist for securing IoT devices. Rather, the crucial point we want to make is that, if a manufacturer is heading into an IoT domain, it should think of itself as a software company. And, any company that takes it upon itself to develop software must also take it upon itself to secure its software. This means committing to taking security engineering seriously, by investing in both educating employees as well as hiring new specialists where needed.

The case for security engineering need not be made from the perspective of civic duty – there are also clear financial arguments supporting such investments. One important example is new legislation underway in Europe. The General Data Protection Regulation (European Parliament, 2016), which will take effect in May 2018, focuses on strengthening and unifying data protection for individuals within the European Union (EU). However, the directive also addresses the exportation of personal data outside the EU. Thus, even some manufacturers outside of the EU will be affected. An in-depth examination of the General Data Protection Regulation is beyond the scope of this article, but it is significant to note that it is broad in scope and covers not only responsibilities and accountability, but also sanctions. Furthermore, the stipulated sanctions are significant. Among them, manufacturers can be fined up to 20M EUR or up to 4% of the annual worldwide turnover of the preceding financial year in case of an enterprise, whichever is greater (European Parliament, 2016: Article 83, paragraph 5–6). Thus, securing IoT devices, which commonly gather data wherever they are in the world, should be made a priority.

Securing IoT devices is the responsibility of a manufacturer’s security engineering team. However, we offer the following initial recommendations to manufacturers:

- Make sure your product's software as well as its operating system can be updated. Make this update automatic, but also make it possible to postpone if the consumer needs to do so. (As the first author of this paper can attest, drones have fallen out of the sky due to unexpected mid-air updates.)

- Try to mitigate human problems. Make it as difficult as possible for the consumer to use their device in an unsafe manner. For instance, passwords as an authentication system are inherently flawed and you should look into adding additional or alternative security tokens. However, if you do use passwords, set up your devices so that default passwords have to be changed when the device is taken into use.

- Learn from the mistakes of the computer industry. One example we brought up earlier is to not leave Telnet enabled. However, there are many other deprecated protocols still in wide use. Close all ports that do not need to be open. Extend this discussion with your security engineering team to include network security.

- Even with a team of security engineers, it is still important to commit both time and resources to security audits and penetration testing. In other words: try to break into your own systems.

- Some vulnerabilities may be found by people from outside of your organization. For this reason, it is important to also have a system in place through which vulnerabilities and bugs can be reported. Offering bug bounties – rewards for finding bugs – may encourage others to find and report vulnerabilities.

There are many, many more things to take into consideration, both regarding security as well as privacy. For this reason, we are not suggesting that manufacturers should follow some external checklist, but rather we urge them to make security engineering a central part of what their company does.

Consumers
IoT vulnerabilities can affect not only a product itself, but also anything from other devices connected to a network to the entire Internet itself. And, again, there are privacy issues, but they are beyond the scope of the current discussion. With the stakes being as high as they are, we encourage consumers to take an active interest in the security of their IoT devices by offering the following recommendations:

- Bear in mind that you are no longer buying washing machines and toasters – you are buying computers that can wash clothes and toast bread. And computers need to be secured. When shopping for an IoT device, be sure to ask about security. Also, check online for known device vulnerabilities before buying.

- When purchasing IoT devices, ask about updates. It is important that you be able to update both the software for the device as well as the operating system that runs it. These updates should preferably be automatic, but with the option to postpone the update if needed.

- Do not buy anything with hard-coded passwords. In other words, if a device uses passwords, it must be possible for you to change the default password.

- Once you have set up your IoT device, always change default passwords immediately.

- Just because a device can connect to a network does not mean that it has to be connected or that that network has to be the Internet. If a connection is required, differentiate between IoT devices that need to be connected to the Internet and those that do not. For instance, if you are installing a security camera, it is likely that you will want to be able to access the feed from the Internet. However, a washing machine, toaster, or any number of other household appliances is likely to be something that does not need to be connected to the Internet. For such appliances, connect them to a local area network, but not to the Internet.

Legislators
There are a number of challenges, both regarding consumer and manufacturer behaviour, that compound the problem of IoT vulnerabilities. We are not entirely hopeful that a greater understanding among manufacturers and consumers of IoT vulnerabilities alone will inspire the necessary actions towards securing the IoT. It seems more likely, if not inevitable, that legislation will be needed to keep IoT vulnerabilities in check. Arguing for legislation has its own problems, and there are certainly examples where legislation has failed. However, it might be that we cannot expect individual manufacturers to invest heavily in IoT security, given that the required investment may hamper their profitability in the name of improving a feature that consumers rarely know to ask about or appreciate. Legislation that makes manufacturers liable for damages caused by the vulnerabilities of their products would force all manufacturers to invest in security engineering, thereby levelling the playing field.

As an example, take home appliances: manufacturer liability for the safety of these devices is already regulated. If your brand-new washing machine short circuits and burns down your house, the manufacturer is liable. Thus, it would seem a small and logical next step to also regulate the security of these devices, making that same manufacturer liable if the damages are of a digital, rather than physical, nature. We do not believe that legislation would need to detail the specifics of how this securing should be accomplished. Merely making manufacturers liable for the cost of not just physical, but also digital faults in their products would ensure a much-needed manufacturer focus on security engineering.

The Internet of (Vulnerable) Things

Mikko Hypponen and Linus Nyman

Conclusion

The IoT revolution is already underway. With its unprecedented number of interconnected computers has come a host of vulnerabilities. These vulnerabilities must be addressed if we are to secure both the future of the IoT as well as a functioning Internet. To achieve this goal, manufacturers will need to put considerable focus on security engineering, policymakers will need to assess the situation to see if legislation is indeed needed to ensure this focus on security engineering takes place, and consumers will need to understand what they can do to minimize the vulnerabilities inherent in their devices.

About the Authors

Mikko Hypponen is Chief Research Officer at F-Secure. He has written about his research for The *New York Times, Wired,* and *Scientific America,* and he has lectured at several universities, among them Stanford, Oxford, and Cambridge. He has been selected as one of the 50 most important people on the web by *PC World Magazine* and was included in the FP Global Thinkers list. He is a member of the board of the Nordic Business Forum and the advisory board of the t2 infosec conference.

Linus Nyman is an Assistant Professor at the Hanken School of Economics in Helsinki, Finland. He has lectured on a range of topics, including corporate strategy and open source software development. His current research focuses on information security and privacy, which are topics he also covers in a blog for the Finnish daily newspaper *Hufvudstadsbladet.* Linus holds a PhD and a Master’s degree, both from the Hanken School of Economics.

References

Abomhara, M., & Køien, G. M. 2015. Cyber Security and the Internet of Things: Vulnerabilities, Threats, Intruders and Attacks. *Journal of Cyber Security,* 4(1): 65–88. http://dx.doi.org/10.13052/jcsm2245-1439.414

Bode, K. 2015. Easily Hacked Tea Kettle Latest to Highlight Pathetic Internet of Things 'Security'. *Techdirt,* October 23, 2015. Accessed April 10, 2017: https://www.techdirt.com/articles/20151015/13551232547/easily-hacked-tea-kettle-latest-to-highlight-pathetic-internet-things-security.shtml

Columbus, L. 2016. Roundup of Internet of Things Forecasts and Market Estimates, 2016. *Forbes,* November 27, 2016. Accessed February 27, 2017: https://www.forbes.com/sites/louiscolumbus/2016/11/27/roundup-of-internet-of-things-forecasts-and-market-estimates-2016/

Corfield, G. 2017. More Brilliant Internet of Things Gadgetry: A £1,300 Mousetrap. *The Register,* February 23. 2017. Accessed April 10, 2017: https://www.theregister.co.uk/2017/02/23/rentokil_1300_pound_iot_mousetrap/

Crook, J. 2016. New Smart Mattress Will Tell You If Your Partner Is Cheating. *TechCrunch,* April 18, 2016. Accessed April 10, 2017: https://techcrunch.com/2016/04/18/new-smart-mattress-will-tell-you-if-your-partner-is-cheating/

Etherington, D., & Conger, K. 2016. Large DDoS Attacks Cause Outages at Twitter, Spotify, and Other Sites. *TechCrunch,* October 21, 2016. Accessed February 27, 2017: https://techcrunch.com/2016/10/21/many-sites-including-twitter-and-spotify-suffering-outage/

European Parliament. 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation).* Brussels: European Parliament. http://eur-lex.europa.eu/eli/reg/2016/679/oj

Franceschi-Biccierai, L 2016a. The Looming Disaster of the Internet of (Hackable) Things. *Motherboard,* November 7, 2016. Accessed April 10, 2017: https://motherboard.vice.com/en_us/article/the-looming-disaster-of-the-internet-of-hackable-things

Franceschi-Biccierai, L 2016b. In the Future, Hackers Will Build Zombie Armies from Internet-Connected Toasters. *Motherboard,* July 5, 2016. Accessed April 10, 2017: https://motherboard.vice.com/en_us/article/in-the-future-hackers-will-build-zombie-armies-from-internet-connected-toasters

Franceschi-Biccierai, L 2016c. The Internet of Things Sucks So Bad Even ‘Amateurish’ Malware Is Enough. *Motherboard,* October 3, 2016. Accessed April 10, 2017: https://motherboard.vice.com/en_us/article/internet-of-things-malware-mirai-ddos

Gartner. 2015. Gartner’s 2015 Hype Cycle for Emerging Technologies Identifies the Computing Innovations That Organizations Should Monitor. *Gartner,* August 18, 2015. Accessed February 27, 2017: https://www.gartner.com/newsroom/id/3114217

The Internet of (Vulnerable) Things

Mikko Hypponen and Linus Nyman

Gershenfeld, N., & Vasseur, J. P. 2014. As Objects Go Online: The Promise (and Pitfalls) of the Internet of Things. *Foreign Affairs,* March/April. Accessed February 27, 2017:
https://www.foreignaffairs.com/articles/2014-02-12/objects-go-online

Graham, J. 2016. Meet the World's First Smart Bra. *USA Today,* January 4, 2016. Accessed April 10, 2017:
https://www.usatoday.com/story/tech/2016/01/04/ces-2016---meet-worlds-first-smart-bra/78247554/

Greenberg, A. 2015. Hackers Can Disable a Sniper Rifle – Or Change Its Target. *Wired,* July 29, 2015. Accessed April 10, 2017:
https://www.wired.com/2015/07/hackers-can-disable-sniper-rifleor-change-target/

Greenberg, A. 2016. The Jeep Hackers Are Back to Prove Car Hacking Can Get Much Worse. *Wired,* August 1, 2016. Accessed April 10, 2017:
https://www.wired.com/2016/08/jeep-hackers-return-high-speed-steering-acceleration-hacks/

Greenberg, A., & Zetter, K. 2015. How the Internet of Things Got Hacked. *Wired,* December 28, 2015. Accessed April 10, 2017:
https://www.wired.com/2015/12/2015-the-year-the-internet-of-things-got-hacked/

Greene, J. 2015. TIM Lecture Series – The Internet of Everything: Fridgebots, Smart Sneakers, and Connected Cars. *Technology Innovation Management Review,* 5(5): 47–49.
http://timreview.ca/article/898

Krebs, B. 2016a. KrebsOnSecurity Hit With Record DDoS. *KrebsOnSecurity,* September 16, 2016. Accessed February 27, 2017:
https://krebsonsecurity.com/2016/09/krebsonsecurity-hit-with-record-ddos/

Krebs, B. 2016b. DDos on DYN Impacts Twitter, Spotify, Reddit. *KrebsOnSecurity,* October 16, 2016. Accessed April 10, 2017:
https://krebsonsecurity.com/2016/10/ddos-on-dyn-impacts-twitter-spotify-reddit/

Krupitzer, C. 2015. Technology Is Improving Our Day-to-Day Lives, from the Boardroom to the… Bathroom? *Thinglogix,* March 12, 2015. Accessed April 10, 2017:
http://www.thinglogix.com/building-the-bathroom-of-the-future-with-iot-technology/

Newman, L. 2016. What We Know about Friday's Massive East Coast Internet Outage. *Wired,* October 21, 2016. Accessed April 10, 2017:
https://www.wired.com/2016/10/internet-outage-ddos-dns-dyn/

Patton, M., Gross, E., Chinn, R., Forbis, S., Walker, L., & Chen, H. 2014. *Uninvited Connections: A Study of Vulnerable Devices on the Internet of Things (IoT).* Paper presented at the IEEE Joint Intelligence and Security Informatics Conference (JISIC), September 24–26, 2014.
http://dx.doi.org/10.1109/JISIC.2014.43

Porup, J. 2016. "Internet of Things" Security Is Hilariously Broken and Getting Worse. *Ars Technica,* January 23, 2016. Accessed April 10, 2017:
https://arstechnica.com/security/2016/01/how-to-search-the-internet-of-things-for-photos-of-sleeping-babies/

Press, G. 2016. Internet of Things By The Numbers: What New Surveys Found. *Forbes,* September 2, 2016. Accessed April 11, 2017:
https://www.forbes.com/sites/gilpress/2016/09/02/internet-of-things-by-the-numbers-what-new-surveys-found/

Schneier, B. 2014. The Internet of Things Is Wildly Insecure – And Often Unpatchable. *Wired,* January 6, 2014. Accessed April 10, 2017:
https://www.wired.com/2014/01/theres-no-good-way-to-patch-the-internet-of-things-and-thats-a-huge-problem/

Schneier, B. 2015. *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World.* New York: Norton & Company.

Schneier, B. 2017. Botnets of Things – The Relentless Push to Add Connectivity to Home Gadgets Is Creating Dangerous Side Effects That Figure to Get Even Worse. *MIT Technology Review,* March/April.

Sims, G. 2015. Samsung Says All Its Products Will Be IoT Enabled within 5 Years. *Android Authority,* January 6, 2015. Accessed February 27, 2017:
http://www.androidauthority.com/samsung-says-products-will-iot-enabled-within-5-years-578576/

Vanhemert, K. 2014. A Toaster that Begs You to Use It: Welcome to the Bizarre Smart Home. *Wired,* March 17, 2014. Accessed April 10, 2017:
https://www.wired.com/2014/03/addicted-products/

Westerlund, M., Leminen, S., & Rajahonka, M. 2014. Designing Business Models for the Internet of Things. *Technology Innovation Management Review,* 4(7): 5–14.
https://timreview.ca/article/807

Citation: Hypponen, M., & Nyman, L. 2017. The Internet of (Vulnerable) Things: On Hypponen's Law, Security Engineering, and IoT Legislation. *Technology Innovation Management Review*, 7(4) 5–11. http://timreview.ca/article/1066

Keywords: cybersecurity, Internet of Things, IoT, smart devices, vulnerability, security engineering, consumers, manufacturers, legislation, Hypponen's law

Big Data and Individual Privacy in the Age of the Internet of Things

Mackenzie Adams

> “*Who could deny that privacy is a jewel? It has always been the mark of privilege, the distinguishing feature of a truly urbane culture. Out of the cave, the tribal teepee, the pueblo, the community fortress, man emerged to build himself a house of his own with a shelter in it for himself and his diversions. Every age has seen it so. The poor might have to huddle together in cities for need's sake, and the frontiersman cling to his neighbors for the sake of protection. But in each civilization, as it advanced, those who could afford it chose the luxury of a withdrawing-place.*”
>
> Phyllis McGinley (1905–1978)
> Pulitzer Prize-winning author and poet

The availability of “big data” and “smart” products are credited with advancing solutions to complex problems in medicine, transportation, and education, among others. However, with big data comes big responsibility. The collection, storage, sharing, and analysis of data are far outpacing individual privacy protections, whether technological or legislative. The Internet of Things (IoT), with its promise to create networks of networks, will magnify individual data privacy threats. Recent data breaches, exposing the personal information of millions of users, provide insight into the vulnerability of personal data. Although seemingly expansive, there are core individual privacy issues that are central to current big data breaches and anticipated IoT threats. This article examines both big data and the IoT using examples of data privacy breaches to illustrate the impact of individual data loss. Furthermore, the article examines the complexity of tackling technological and legislative challenges in protecting individual privacy. It concludes by summarizing these issues in terms of the future implications of the IoT and the loss of privacy.

Introduction

Across most domains, societal functioning has become increasingly dependent on information and communication technology, as well as the management of massive data streaming through physical and virtual environments. The generation of this extensive data, formal or informal in structure, has led to its referral as “big data”, a nomenclature pointing to not only sheer size, but also to the speed with which it is generated and the complexity in organizing and analyzing it (Berman, 2013; Chen et al., 2014). Big data has emerged as an area of significant interest in research and applications for organizations dealing with or anticipating an overwhelming flow of data. Individual privacy regarding big data has especially taken hold as a central issue affecting different technology areas as connectivity and information sharing have far outpaced data protection efforts (Perera et al., 2015).

Widely publicized breaches of large databases exposed significant and escalating threats to individual privacy and control over personal data. In 2005, a security breach of an American health insurance company, Anthem, led to the theft of personal information of more than 78 million customers (Mathews, 2015). The information included names, dates of birth, social security numbers, and income data, all of which were likely sold in underground markets. The total number of affected individuals and the sensitive nature of large data breaches are alarming; they also point to an urgent need to the convergence of technology, legislation, user policies, and awareness in protecting privacy.

Big data and individual privacy protection are further complicated by the evolution of networks of networks, also referred to as the Internet of Things (IoT). This new paradigm promises to enable existing and future devices to be connected to local and virtual networks and, eventually, communicate autonomously with these networks and other devices for functions such as gathering and analyzing data (Borgohain et al., 2015). For instance, new applications are enabling users to check the status of their home appliances from their smartphones, monitor private property, and synchronize their devices while increasing the likelihood of exposing the large amount of data collected and stored in these devices and networks to other individuals and entities.

According to Russo and colleagues (2015), by 2020, there will be over 200 billion sensor devices that are interconnected. These sensors will be found in home electronic systems, health monitoring equipment, cars, and smartphones. Their economic impact will also be tremendous, according to the authors who estimate that, by 2025, their market will be worth approximately $3 trillion per year. As the surface area of data expands exponentially through the IoT, the implications of individual privacy threats of this pervasive interconnectivity are immense. Current breaches of large databases and their impact provide insights into how the future of big data and the IoT is shaped. It becomes of significant importance to explore how the collection, storage, sharing, and analysis of big data can be complex and multifaceted and how it can bridge the worlds of technology and application development, privacy legislation, and consumer/user privacy protection processes.

This article examines the implications of compromised individual privacy in the age of the IOT as it relates to big data. First, it provides definitions and descriptions of the widely used terms “big data” and the “IoT”. It clarifies the parameters used by researchers in studying and writing about both phenomena, as well as touches upon vulnerability that expose the privacy of individuals’ data to unauthorized access, loss, or theft. Next, it examines the extent to which recent big data breaches have exposed the vulnerability of personal data. The examples illustrate the different pathways and impact of individual data loss. Then, the article places issues and challenges of data privacy loss into the context of the age of the IoT, and it emphasizes the fundamental complexity of the IoT and the how it is likely to present further technological, legislative, and user experience challenges to protecting individual privacy. Finally, the article integrates and summarizes the previous sections by examining opportunities in security and individual privacy protection in the age of the IoT.

The underlying assumption of the article is that the collection of data from IoT devices and customization based on the collected data create vulnerabilities in individual data privacy. As a framework to guide the discussion, Figure 1 provides an overview of individual privacy when big data is examined in the age of IoT. Individual privacy is threatened when data is collected and a data breach can expose an individual’s private data; it is also threatened when companies and individuals, under the pretext of assumed consent to provide a custom experience, use the collected data. The roles of technological and legislative solutions in protecting individual data privacy continue to change and evolve.

Big Data, IoT, and Data Privacy

Big data, as a concept, has been around for two decades since being used by Cox and Ellsworth (1997). While initially referring to extensive volumes of scientific data, big data has since been defined in a number of ways. Boyd and Crawford (2012) argue that it “is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets”, whereas Hashem and colleagues (2015) propose that big data has three characteristics: i) numerous, ii) cannot be cat-

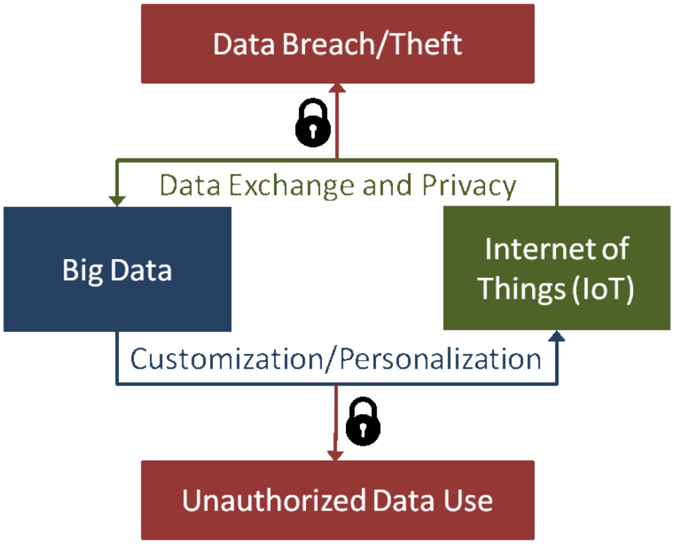


Figure 1. A framework for big data and individual privacy

egorized into regular relational databases, and iii) generated, captured, and processed rapidly. Chen and Lin (2014), on the other hand, define big data as "the exponential growth and wide availability of digital data that are difficult or even impossible to be managed and analyzed using conventional software tools and technologies".

The most commonly known definition was suggested by IBM (Malik, 2013; Schroeck et al., 2012), which proposes that big data is characterized by any or all of the following three attributes: volume, variety, and velocity. *Volume* reflects the tremendous amounts of data created from a number of sources and across different platforms such as mobile devices and applications and smart grids, as well as social media such as Facebook. The sheer volume of big data is likely to increase substantially as IoT-enabled technology will continue to be designed to generate data from multiple devices and sources. *Variety* refers to the nature of data generated. For instance, structured data from geographic information systems as well as unstructured data from websites are found in numerous formats. *Velocity* reflects the speed with which data is not only generated from a myriad of sources, but the frequency of data capture, analysis, and the application of information in decision making. Hashem and colleagues (2015) have added a fourth "v" to the IBM definition, "value", noting that it is the "most important aspect of big data; it refers to the process of discovering huge hidden values from large datasets with various types and rapid generation". Thus, *value* refers to the actual use of the data collected. Physical devices or sensors may not, by themselves, provide data that can be used for predictive modelling in medicine or retail, for instance. However, multiple devices and sensors can provide data that, when aggregated, provides valuable information upon analysis.

Big data, therefore, is likely about the above four attributes and their scaling to ever greater numbers of devices, infrastructures, and networks. At its core, big data describes the wide availability of data in digital form, with a concomitant presence of data mining and knowledge-generation capability across numerous networks.

Mining big data

The collection and storage of large volumes of data has held the promise of data-driven discovery in diverse fields including scientific research, healthcare, industry, manufacturing and education (Chen et al., 2015; Malik, 2014). Massive volumes coupled with wider availability aimed to fulfill this promise through the development of data exploration and mining technologies. The purpose of data mining, therefore, is to uncover useful and novel information from data stored in large databases, thereby being predictive or descriptive. This is an especially important development in fields reliant upon large data for making those predictions to be generalized across populations such as medicine and commerce. The data mining process, in general, involves several major steps whereby data is cleaned, transformed, and mined for information.

Big data and the use of machine learning algorithms have become inextricably linked with data mining recently. A main reason is that datasets have grown larger and more complex, and traditional learning methods of managing such volumes while extracting useful data have fallen short. Furthermore, while the volume of data has increased, its quality has remained inconsistent; data mining efforts face low quality, multi-form data across numerous applications and systems, and are further complicated by the lack of effective security solutions to share such data. As noted by Shukla (2015):

> *"I use the term big data a bit too generically to include machine learning and data mining even when the data is not necessarily 'big'. Especially when the Internet of Things becomes a reality in improving the lives of people, improving quality of automation systems, and improving transportation system performance, machine learning and data mining will be ready to deliver technologies, algorithms, and possibly products that can be directly used to make those systems perform in the most optimal fashion, adapting to changing situations, and securing the system against hackers who would certainly want to disrupt such systems or try to breach privacy of people who will be connected to such networks."*

Data mining for effective decision making may seem innocuous from the perspective of private data exposure. Aggregate forms of data, such as those collected by search engine programs or presented in census information, are expected to remove key pieces of identifying information while retaining others for the purpose of analysis (Boyd & Crawford, 2012; Liu, 2014). For instance, census data collected may aggregate ages to arrive at descriptive statistics for age groups, but will be expected to not provide access individual identifying information such as names and addresses. However, these expectations are outside the control of individuals whose data may be stored, transferred, shared, and analyzed by different individuals and organizations. As both data volume and data mining interest increase in the IoT paradigm, the issue of privacy becomes more urgent.

Big Data and Individual Privacy in the Age of the Internet of Things

Mackenzie Adams

Internet of Things (IoT): Current paradigm and anticipated reality

From a review of recent literature, it is apparent that the IoT encompasses an understanding of how networks of networks will connect devices, infrastructure, and systems, among others, through a new Internet. The review shows that the IoT is referred to by researchers and practitioners as “a vision”, “a new paradigm”, “an area of research”, “an emerging global Internet based information architecture”, “next step evolution of our today Internet”, “a growing technology”, and “a new form of computation”. Perera and colleagues. (2015) define the IoT as a “network of networks, in which, typically, a massive number of objects, things, sensors, devices are connected through the information and communications infrastructure to provide value-added services”.

A comprehensive definition of the IoT is also presented by Russo and colleagues (2015), who state that:

> *“The Internet of Things (IoT) is an integrated part of the Future Internet and can be defined as a dynamic global network infrastructure with self-configuring capabilities based on standard and inter-operable communication protocols, where physical and virtual things have identities, physical attributes and virtual personalities; they use intelligent interfaces and are seamlessly integrated into the information network.”*

The IoT promises unprecedented advancements across knowledge-based industries and fields. According to a review of literature on the IoT (Russo et al., 2015) from its earliest conceptions in the late 1980s to 2015, numerous future characteristics behind these advancements are proposed by researchers as follows:

- Evolution in communication, not only human–human and human–things, but things–things as well, reflecting an increasing role of autonomous communication among devices and artificial intelligence research and application.

- Optimization of energy consumption through network infrastructures and remotely monitored systems designed to reduce consumption. Smart homes are an example whereby devices can be programmed to autonomously communicate and can affect such things as temperature settings and electricity consumption.

- Wider opportunities to develop technologies and tools through the creation of Internet-connected devices.

- Greater role in development of technologies in medicine, critical infrastructures, and smart cities. Recent advances in continuous patient monitoring, including in-hospital and out-of-hospital applications are strong examples of such technologies.

On a wider, societal scale, IoT applications are numerous and wide-ranging given that they are used in commercial, environmental, and critical infrastructure settings (Chen et al., 2014). It is expected that, with an increased capability in analyzing large data, high-quality information will guide such functions as monitoring air quality and pollution indices, as well as monitoring food as it is transported across the globe. The agricultural industry can exploit in-ground sensors and irrigation-control software to automate its soil management, while reducing costs associated with inclement conditions (Russo et al., 2015). Commercial applications have noted the ever-increasing role of supply chain management and logistics, both of which are made more efficient and cost-effective when connected devices are programmed to provide basic decision-making capability.

In summary, the IoT will allow billions of objects, such as mobile devices, and virtual environments to exchange data. With machine learning, devices and environments may exchange such data autonomously while extracting meaningful data. However, the IoT – by definition – is complex and covers extensive data landscapes, structures, and contexts. This complexity has serious implications in securing information flowing from individuals’ devices to the networks of the IoT. To further complicate the exposure of private data, cloud computing environments essentially upload the ’minute details of one’s life to virtual environments that are targets for privacy breaches (Maras, 2015; Matzner, 2014; Perera et al., 2015). The IoT is a developing target for interconnectivity of devices and environments in a network of networks. The potential entry points and vulnerabilities to data privacy breaches are also developing, and a key question is whether security measures can be concomitantly interoperable and scalable. However, breaches of large datasets are a reality, and recent years have shown how vulnerable individual data is to loss of control, theft, and exploitation.

Privacy Loss and Big Data Breaches

Privacy of individual data is expectedly complex and multi-faceted, extending across technological, legal, commercial, and financial domains (Punagin & Arya, 2015). The loss of personal information to unauthorized and illegal means did not start with the Internet; individuals were likely to lose their financial information such as credit card statements or social insurance numbers from thieves rummaging through personal effects or property. The widespread digitization of everyday living, from financial transactions to personal communication, to business dealings, however, has exposed individual information to unauthorized access to entities from across the globe (Bekara, 2014). In the process, it has prompted a revisiting of privacy threats and an examination of individual privacy and control of data generated by our activities as a right deserving of user and legal protections. It remains that the right to the massive data collected currently through databases – which are expected to be interconnected, sometimes autonomously, through the IoT – has legal frameworks and privacy-enhancing technologies but they are lagging to provide adequate protections (Han et al., 2014; Maras, 2015).

Some examples of big data collection may seem mundane. Currently, most smartphones are enabled with location sensors, providing real-time data to be collected on an individual's whereabouts and activities (Rghioui, et al., 2015). As more devices are enabled to provide similar information, we observe that cars also provide data on location, while household efficiency and security protection are connected to handheld devices. Taken together, the information from disparate devices provide extensive information on individual and behavioural patterns, which is a privacy concern (Schroeck et al., 2012; van de Pas & van Bussel, 2015). This situation is similar to the collection of browsing history and purchasing behaviour used to tailor online activities to an individual. However, they are also similar in exposing individuals to the loss of their information.

Big data collection and mining are also promising to transform the quality of individuals' lives in innumerable ways. In healthcare, for instance, health information collection is now enabled in many everyday devices such as iPhones or FitBits, providing continuous data collection of key health behaviour, a function reserved in the past through medical intervention to a limited number of people (Suciu et al. 2015; Tsai et al., 2014). Abinaya, Kumar, and Swathika (2015) examined the application of the IoT in devising an information system based on the ontology method. The researchers explored a system that aimed to connect emergency medical services with hospital-based services.

The implications of this data collection and storage, and the ability to provide real-time analysis and provision to healthcare providers, represent a revolutionary advancement in health monitoring and preventive care (Abinaya et al., 2015). With an increase in big data analytics and technology, the large, raw health data collected from these and other devices can provide valuable information about the individual's health, as well as population-level information that previously would have only been available through formal, large studies. Once again, however, privacy risks are inherent in the collection, storage, and exchange of this data. Individuals may lose control of who views their information, which has the potential to result in exposure of health conditions and practices, but may also have ramifications for employment and health insurance (Borgohain et al., 2015; Krotoszynski, 2015).

High profile data breaches, especially of businesses, often dominate media coverage of data security compromises because they often involve the information of numerous clients and customers. A data breach is said to have occurred when individuals' data has been subjected to unauthorized access, resulting in the exposure of confidential, protected, or sensitive information. The personal, financial, and legal impact of data breaches can be tremendous (Sen & Borle, 2015). Individuals whose information is stolen or accessed can suffer identity and financial losses, and have sensitive information such as health conditions or personal behaviour scrutinized and exposed. Organizations that are breached are also likely to suffer financial and proprietary information losses, as well as reputation compromises. Organizations that collect extensive personal data from their customers, such as healthcare institutions and banks, are particularly vulnerable to such losses. According to the Ponemon Institute Report (2014), the impact of data breaches' on individuals, mostly linked to identity thefts, are implicated in a loss of $16 billion approximately from nearly 13 million individuals. The average cost per incident was estimated to be nearly $6 million for organizations in the United States. The report also cites that identity theft is the dominant consumer fraud complaint to the United States Federal Trade Commission (FTC).

Financial and private institutions

A number of illustrative cases of big data breaches in recent years have shed light on the nature and impact of individual data security compromises. In the previously mentioned Anthem health insurance company breach, approximately 78 million people had their company private records illegally accessed (Mathews, 2015). Breached data included their identifying numbers along with names, dates of birth, and social security numbers, as well as income data. The nature of the data stolen reflects the high risks of identity-theft schemes. The hacking of a large database of JPMorgan Chase bank affected a similar number of individuals, approximately 76 million (Silver-Greenberg et al., 2014). The database hackers gained access to applications that were run on the bank's computers where they were able to exploit a known vulnerability. The hackers were able to access personal details such as names, addresses, and phone numbers, although the company had released a statement that other personal data such as dates of birth were not included in the hacked databases.

Given the wide use of social media and networking sites, it was inevitable that a large data breach would occur. The Canadian owned social dating site, Ashley Madison, was hacked in 2015, exposing the company's internal servers, company bank account data, and staff salary information (Solomon, 2015).

Public institution breaches

Although they are attractive targets for big data breaches, financial institutions are not the only organizations that are targeted for malicious access. Approximately 191 million American voters' personal information was exposed on the open Internet due to an incorrectly configured database (Finkle & Volz, 2015). While not considered a malicious act, it is, nonetheless, a data breach that exposed the personal details such as name and address, as well as party affiliations of voters in all 50 States and Washington, DC. Another governmental body exposed the individual private data of millions of American military veterans when a breach occurred at the National Archives and Records Administration (Singel, 2009). The breach was traced back to a defective hard drive that the organization had sent to the external vendor for repair. However, it was later discovered that the data recorded in the drive was not destroyed before being sent to the vendor.

Those seeking illegal access to data are, at times, motivated by nation-state purposes. An example of a public institution breach through such a purpose is the hacking of the Office of Personnel Management (OPM) in the United States by the Chinese state (Nakashima, 2015). The organization informed approximately 4 million current and former federal employees that their personal data had been accessed illegally. Representing the biggest data breach of federal employees in recent history, the OPM breach exposed personal identifying information such as social security numbers, human resources' related information, and job assignments.

Critical infrastructures are also a target for big data breaches. The San Francisco Public Utilities Commission warned approximately 180,000 thousand of its customers that a data breach had exposed their personal information to illegal access (Mills, 2011). Specifically, customers' account numbers and personal identifying information such as names and addresses were breached. According to the organization, the breach occurred when an unsecured server was infected with viruses through an open port.

Individual data loss impact and protection

The above-illustrated cases of big data breaches provide insight into both the vulnerability of personal data and the impact of its loss. Organizations must work to secure personal data by ensuring that only information that is required is collected from users and customers. Ensuring that only required information is collected will force both individuals and organizations to realize that data has to be protected, and the less personal/sensitive data collected, the less likely that it is breached (Maras, 2015). Furthermore, to safeguard personal information, it is crucial that storage and transportation processes are embedded with security measures. The above examples of inadvertent data breaches show that carelessness, poor follow-through, and lack of accountability can be just as harmful as intentional hacking or malicious behaviour.

Organizations should also consider effective and periodic ways to discard personal information collected from individuals, especially when that information is no longer required in its raw forms. To reduce the risk of unused servers becoming the target of data loss, users and organizations should be diligent in pursuing strict and accountable processes for discarding data. As explored in this article, there are important implications of inconsistent data management and handling processes that will surely be magnified in IoT environments (Maras, 2015; Samani et al., 2015). When the absolute volume of data exchanged increases exponentially in such environments, even the most diligent of systems can "lose track" of personal information, especially as data is streamed from new devices and objects.

It is also important to consider that the public–private sphere of policies and protections are at times blurred in the context of data exchanges (Schroeck et al., 2012; van de Pas & van Bussel, 2015). For example, policies to limit data collection in public institutions may not exist in private organizations. Governments are likely limited in how they can impose data protection measures in the "private sphere". Standardization of measures, even within public institutions, is a challenge due to a potential impact of data exchange limits. Potential employees or health insurance seekers are expected to provide their personal information. If they do not, they may not be insured or considered for employment. The same can be said for everyday aspects of life including securing loans, buying or renting places of residence, and even enrolling in colleges and universities. Thus, individuals in society cannot opt out of disclosing their personal information to private and public entities, but such disclosure comes with the risk that their private information may be exposed in a data breach.

Individual Privacy Issues and Challenges in the IoT

As more IoT-enabled devices and systems are created, more individual data privacy issues and challenges emerge, especially as big data analytics and technology are positioned to search for value in this data. It is generally insightful to examine "on-the-ground" applications of the IoT against emerging privacy concerns. For instance, Rghioui and colleagues (2015) examined the lack of consideration of data security and privacy in the IoT-based wireless body area network (WBAN). Specifically, the researchers reviewed various devices that are now attached to patients physically to monitor health outputs such as cardiac function. These devices have allowed patients to become more mobile while continuously monitored by their healthcare providers and transmitting data through the WBAN. Rghioui and colleagues (2015), however, found that, despite the tremendous advancement in health monitoring offered by these devices, the WBAN networks were largely open to outside access with external IP hosts, which could compromise data integrity, disrupt communication between the mobile devices and the networks, and expose personal health information to unauthorized individuals.

Rghioui and colleagues (2015) proposed a number of solutions for the management of security keys through encryption, which would consider patient mobility and a device's resource constraints. The solutions they proposed address a number of important factors in addressing IoT data-privacy issues, namely, data integrity, scalability, mobility, and key connectivity. Data integrity is an especially important factor whereby encryption keys ensure that no unauthorized access occurs in the transfer of information among devices and the networks. Scalability is also important given that a key challenge in security measures in the IoT is whether a network can remain stable as more devices are added to it. Although their proposed solutions in managing privacy concerns in a healthcare setting are technologically focused, their paper sheds light on overarching issues in securing the integrity and access to large volumes of data in an IoT environment, while continuing to scale up the technology to serve more patients in greater health monitoring functions.

In addition to healthcare, smart grids are an area of exploring big data privacy issues and challenges in the IoT. Bekara (2014) examined security as a determining factor in the expanded application of the IoT and smart grids. iIn a number of IoT-based smart infrastructure contexts, such as homes, cars, and appliances, inherent data privacy and security issues include: impersonation/identity spoofing; eavesdropping; data tampering; authorization and control access issues; privacy issues; compromising and malicious code; and availability and denial-of-service issues and cyber-attacks. Thus, individual data privacy in an IoT-based smart grid is largely compromised through exposure of personal information to unauthorized access, especially in the context of device-device and device-network communication. Similarly, Bekara (2014) has highlighted privacy and security challenges related to scalability; mobility; deployment over large areas; legacy systems; constrained resources; heterogeneity in implemented protocols and communication stacks; interoperability; bootstrapping; trust management; and latency or time constraints.

Other researchers examining IoT-enabled technologies, especially those affecting individuals and households, note similar data privacy and security challenges. Punagin and Arya (2015) also explore the various opportunities presented through IoT-based technologies such as healthcare, mobility, smart grids, law enforcement, and e-commerce. The researchers note a number of similar privacy and security challenges as well, such as identity/sensitive attribute disclosure. As noted earlier, it is expected that big data is an aggregate of individual data, and that various methods of de-anonymizing individual-level data will be available. However, published data may be susceptible to external linkage attacks where hackers and other attackers can link the publicly available data to the de-anonymized one. Narayanan and Shmatikov (2008) were able to de-anonymize a Netflix data set, linking it to individual user profile data from an entertainment repository website, while Sweeney (2002) de-anonymized a hospital's anonymized health records by linking the data set with publicly available information.

Punagin and Arya (2015) also include automated recommendations. Big data may expose a person's behavioural patterns (websites visited, pages clicked, etc.) on their social networking site. The automated recommendation may be a data breach if the person has not provided consent. Finally, the researchers list predictive analysis as a security challenge. According to the authors, retailers could use big data analytics to conduct regression analysis on individual purchase habits and patterns, and use them to make predictions about future behaviour. Although this approach may be used widely on a population level, at an individual level, it is a data privacy compromise, especially when consent is absent.

Hashem and colleagues (2015) examined data privacy and security challenges in cloud computing applications. Cloud computing refers to distributed data-processing platforms, and it is one of the building blocks of IoT-based technologies. The authors note, "big data utilizes distributed storage technology based on cloud computing rather than local storage attached to a computer or electronic device. Big data evaluation is driven by fast-growing cloud-based applications developed using virtualized technologies" (Hashem et al., 2015). Given this, privacy and security challenges include big data mining and analytics, which access personal data to create information such as location-based services and recommendations. The authors argue that this use of individual data exposes individual privacy to profiling, loss of control, and theft. The authors further note that control over individual data falls under rules of transparency and accountability that exist between users and organizations, and these rules must be clarified in cloud computing given the high chance of individual privacy compromise.

In their analysis of individual data privacy in the era of the IoT, Perera and colleagues (2015) focus on the inherent assumptions and understandings of which users must be aware when connecting to the Internet with their devices. For instance, the authors note that, when individuals use free online services, such as Facebook and email, they must be aware that they are signing on to become sources of business data. This data is likely used by the service owners to improve services; however, it may also be used to conduct predictive analyses or may be given to affiliate businesses and organizations. Consent may or may not be sought for these actions. Perera 'and colleagues (2015) predict that consumers may find themselves weighing the "free" aspect of online services against their privacy protection in connecting to IoT-enabled technologies. This is especially the case with these technologies continuing to gather more intimate personal information such as health metrics and daily living behaviours. If not paying outright for privacy protections, individuals may opt to limit how their data is used in exchange for continuing to use free services.

Individual Data Privacy Protection in the IoT

Earlier, we noted that big data and IoT-enabled technologies have outpaced the development of legal and user-privacy protection frameworks'. Weber (2015) argues that today's IoT devices are designed to minimize the likelihood that data transmitted across devices and networks will be at risk for tampering and interception. However, he notes that existing protocols and compression technologies for the movement of large volumes of data are limited. Furthermore, the technological limitation is coupled with legislative ones that have not caught up with fast advancements in the field. There is little argument that privacy is considered a right, and individual user protections are necessary to safeguard this right (Maras, 2015). Legal data-protection laws and privacy laws are limited, however, by the type of data created, collected, transmitted, and exchanged. For instance, the European Data Protection Directive (DPD) legislates data if it is deemed private (Weber, 2015).

Privacy definition and legislation

The definition of privacy is understandably diverse and broad. In 1968, Westin defined "information privacy" as "the right to select what personal information about me is known to what people". The definition is dated but has a core value of "right" in controlling individual information disclosed to others – a value that is significant even in the era of the IoT. Ziegoldorf, Morchon, and Wehrle (2014) proposed an IoT-relevant definition of privacy that is reflective of current technological innovation and data exchange. The authors' definition of privacy in the IoT is the "guarantee to the subject for 1) awareness of privacy risks imposed by smart things and services surrounding the data subject; 2) individual control over the collection and processing of personal information by the surrounding smart things; and 3) awareness and control of subsequent use and dissemination of personal information by those entities to any entity outside the subject's personal control sphere" (Ziegoldorf et al., 2014).

Privacy legislation aims to provide a balancing force against business and commercial enterprises' ever-increasing chase of data that services market and advertising needs. With appropriate legislation, individual privacy protections place the values of personal information control and use as prime values in this balancing act. The 1948 Universal Declaration of Human Rights recognized privacy as a fundamental human right, while most countries' constitutional rights include privacy (Ziegoldorf et al., 2015). The United States passed the first known legislation on information privacy more than 40 years ago through the 1974 US Privacy Act, whereby fair information practices (FIPs) were established. The FIPs were developed to hold a number of core values regarding individual information including "the principles of notice, consent, individual access and control, data minimization, purposeful use, adequate security, and accountability" (Ziegoldorf et al., 2015).

Regardless of the core values and principles underlying current existing privacy legislations, there are fundamental challenges in the era of the IoT, as pointed out by several researchers (Krotoszynski, 2015; Maras, 2015; van de Pas & van Bussel, 2015; Ziegoldorf et al., 2015). One important challenge is the definition of "personal" in a number of concepts such as "personally identifiable information". Attributes such as date of birth and financial information as identifying attributes in definitions may vary by legislation or jurisdiction. This variability makes it a challenge to have a single privacy definition that could apply across different technologies and applications in the IoT that are developed and managed by different entities.

A second challenge identified by researchers is how legal frameworks and legislations lag behind applications going live and being used by millions worldwide. Ziegoldorf and colleagues (2015) note that the European Commission passed a law against the tracking of web users in 2011; this legislation comes nearly 20 years after users starting browsing the web. In a similar vein, IoT-enabled technology is developing at a much faster rate than legislation could and should. It remains that many jurisdictions have not legislated the sale of user data on websites that offer their services free, such as email. Thus, users are likely to receive promotional and other marketing information once they have registered to use a free site. With IoT technologies, Ziegoldorf and colleagues (2015) argue that it is unclear whether "personal" information in the future will include readouts from health monitoring devices or home smart meter readings.

A third challenge for privacy legislation in the IoT is unique to the paradigm: the speed with which data is exchanged and the volume of data involved both make it unlikely that data privacy breaches will even be known to individuals. Unlike previous data breaches that could be linked often to financial fraud or identity theft directly, and thus individuals were made aware of them through their credit reports and financial statements, the loss of personal information from multiple devices is more insidious. One can lose data privacy aimed to individualize advertising without a physical loss of assets or exposure of private data in a public platform. For example, output from a medical device could be used by others to tune their advertising, but it still reflects loss of personal information.

Privacy protecting solutions in the IoT

Protecting individual data privacy in the IoT will bridge legislative and technological solutions, in addition to addressing social, cultural, and political factors. The purpose behind any data privacy protection solution will be compliance; however, there are a number of challenges that impede such compliance. If system development does not integrate sufficient privacy-protecting capabilities, expanding them upon and beyond deployment is often costly, unwieldy, or not possible (van de Pas & van Bussel, 2015). Similarly, when protection solutions include policy and user documentation that are vague in language, inadequate in scope, and non-enforceable across applications and systems in an IoT environment, compliance is also affected. Spiekermann and Cranor (2009) provide a framework whereby privacy can be protected through two major routes: privacy-by-architecture and privacy-by-policy.

Privacy-by-architecture aims to incorporate privacy-preserving functionalities into the earliest stages of system development. For instance, while gathering system requirements, the engineers and developers will aim to build capabilities that minimize the collection of personal data or provide anonymization functionality during the information lifecycle. Privacy-enhancing technologies use this process in their development. Privacy-by-policy, on the other hand, holds "notice and choice" as a central value in developing privacy-protecting policies. Spiekermann and Cranor (2009) note that, despite the expedience of this approach, it has multiple issues that fall short of providing effective protection of individual data. Specifically, organizations such as companies, service providers, and data-collecting governmental bodies can readily draft privacy policies that maximize their access to individuals' personal information while writing the protection components in vague language that is difficult to understand (Maras, 2015). When these entities incorporate language that is designed to provide defense against future lawsuits, privacy-protecting policies become even more incomprehensible to the average user (Krotoszynski, 2015; Samani et al., 2015). For privacy-by-policy to provide effective protection solutions to individual data in the era of IoT, it will address these challenges through user-controlled language and parameters, as we will note shortly.

Privacy-by-architecture
Currently, there are a number of privacy-enabling technologies that are deployed to provide some protections. Suciu and colleagues (2015) looked at how to secure e-health architecture through a search-based application, CloudView. Specifically, they noted how cloud middleware received data from heterogeneous devices and integrated data from healthcare platforms, which at times compromised the security of user information. Their proposed search-based application protects this user information by ensuring that data is stored and processed as close as possible, in both space and time, to its location of creation and consumption. The researchers also supported non-functional requirements in the solution such as reliability and security through well-designed integration of physical resources and remote devices, thus “things” and gateways. Finally, the application ensured the distribution of on-the-spot inferred content, instead of raw data. This quality of the solution reflected resource efficiency and scalability of the system so that more IoT-enabled devices and objects can be added.

In their overview of security protecting solutions in cloud applications, Hashem and colleagues (2015) note a number of approaches, including the development of a reconstruction algorithm for privacy-preserving data mining. They also note that a privacy-preserving layer can be applied over a MapReduce framework to reduce risks to privacy caused by data indexing. The privacy-preserving layer makes certain that data privacy is preserved before it is further processed, while ensuring that other data processing applications can be integrated. Given that many privacy-protection solutions are resource intensive and, thus, cannot be scaled in IoT environments, the researchers propose a solution that is an “upper bound privacy leakage constraint-based” approach. To make encryption of data feasible in cloud computing, the solution helps identify which intermediate datasets should be encrypted rather than encrypting all. The benefit is that protection of data can be effective without incurring the cost and time of encrypting all datasets in various states of cleaning, transformation and analysis.

Henze and colleagues (2016) also provide privacy-protection solutions for cloud-based IoT technologies and applications. The authors do so by allowing users to enforce their privacy requirements before their sensitive data is uploaded to the cloud. The solution also enables developers of cloud services’ to integrate this privacy functionality into existing IoT-enabled devices. The core requirements of a system that integrates the IoT and cloud computing in privacy-critical application areas are as follows (Henze et al., 2016):

1. *Data security* ensures that data access is controllable by the owner of the data. Security design and mechanisms have to be robust and flexible enough to allow owners to change their mind about access in the future.

2. *Transparency by design,* on the other hand, and as recommended by van de Pas and van Bussel (2015), ensures that data-usage documentation is incorporated into the design and implementation of a cloud service so that users have transparency regarding how their information will be accessed and by whom.

3. Similarly, *privacy-aware development* ensures efficiency in enhancing privacy-protection capabilities by supporting these functionalities early in the development process.

4. *User-controlled* data use and handling shift the control of data access and use to the individual end user rather than the developer or service provider.

5. *Adaptable user control* allows for differential expertise in these end users to tailor data access and use control to their needs in the future.

Privacy-by-policy
Protecting individual data privacy through policy is common practice that is, similar to legislation, likely to fall short in IoT contexts. Both the lack of clarity in language and poor classification of “private” or “personal” information across applications and systems are important factors. However, there are steps towards privacy protection solutions through policy that address these factors. Lu and colleagues (2015) propose an attribute-based privacy information security classification (PISC) model that classifies information into categories based on the degree of security and privacy. Each classification is designed to have a security goal that determines the nature of encryption, access control, and a time limit for access.

Punagin and Arya (2015) argue that privacy protection can often come at the expense of utility and access to online services, with a resultant restriction in the information provided as more security measures are implemented. They propose that users should be in control of how much of their private information they are willing to share with others, at the risk of exposure, to achieve better utility. Thus, "data collection and usage mining becomes transparent and users understand what data is being mined and how it is used, they may be willing to share their personal information with increased confidence" (Punagin & Arya, 2015).

Other data privacy researchers propose that policy-carrying data is an effective solution for incorporating user control into the development of data-protection policies. Padgeta and Vasconcelosb (2015) acknowledge that the "who", "when", and "how" concerns of data access must be captured in policies to protect data privacy. They propose that a way to capture the wording and manner of access controls over data, and the ability to link that with clarity with the data through what they term "policy-carrying data" (PCD). According to the authors, the PCD sets parameters for the transmission, storage, use, and disposal permissions. The formalized process would provide very specific instructions to how pieces of data can be used and by whom. The following is an example of a PCD proposed by the authors:

"Lab managers can access 500 records of my data. If an interested party requested 1,000 records, the server would (i) check the credentials of the requester (who needs to be registered); (ii) grant access to 500 records (a message would provide reasons for not providing the 1,000 records); (iii) update the record of that requester with respect to that PCD. Further requests from the same party would be rejected with a suitable justification." (Padgeta & Vasconcelosb, 2015)

There are several qualities of this PCD that address privacy policy challenges presented above. One quality is the specificity of the data use and control. There is an upper limit, with a provision on how to handle more requests for data. The PCD also has clear language that is controlled by the data owner. Rather than vague, often standard, language about the use of data, it provides clear parameters and consequences for requests beyond those parameters. More importantly, it places transparency as a core factor in communicating data access and control wishes.

Saroiu, Wolman, and Agarwal (2015) also propose the use of PCD to provide individual data-privacy protections in cloud-based applications. The authors argue that, instead of expensive and difficult-to-implement technological solutions, individuals should use a simpler approach before uploading to a cloud environment any data they deem private. Their form of PCD, as a terms-of-service document similar to the one used by sites and service providers already, will allow data owners to be the ones to dictate how their data will be used. The main purpose of the PCD proposed is to bind the user's data to the policy parameters and conditions of use. Therefore, an individual can be explicit in opting out of (or into) some data uses or in setting time/volume limits as proposed by others.

It is interesting to note that the proposal by Saroiu and colleagues (2015) uses encryption in a novel way. It compels the cloud services' providers to be compliant with the PCD that the data owner attaches to the data. It does so by using ciphertext-based attribute-based encryption (CPABE). Following their reading of the PCD, the service providers must build a number of attributes that are compliant with the policy parameters and conditions. If the attributes are not compliant, the decryption fails and the data is not available in the environment. Similar to the Padgeta and Vasconcelosb (2015) approach, this PCD places data owner control as a core value in creating a policy-driven data protection solution.

Conclusion

Protecting personal data in the era of big data and the IoT requires a multi-faceted approach that places data owner control as a core value of its solutions. Individuals must be not only aware of the data they generate and share across devices and platforms, but they must also understand the security risks and implications of a breach. Whether technology or policy, or a combination, is used to protect individual data, it must be done with users controlling who accesses their information and in what manner. And, importantly, data owners should not be penalized for accessing the advantages of an increasingly connected, data-rich world of information and communication technology with an increased risk of privacy loss and exploitation.

Big Data and Individual Privacy in the Age of the Internet of Things

Mackenzie Adams

About the Author

Mackenzie Adams is Co-Founder and Creative Director at SOMANDA Inc., and she is a recent graduate of the Technology Innovation Management (TIM) program at Carleton University in Ottawa, Canada. As an avid learner and serial entrepreneur, Mackenzie is always seeking new challenges to continue evolving and expanding her interests, knowledge base, and skills. Her interests span the fields of artificial intelligence, quantum computing, EdTech, and FinTech. Her passion is to find and cultivate the next generation of innovators in underserved communities.

References

Abinaya, V., Kumar, V., & Swathika, K. 2015. Ontology Based Public Healthcare System in Internet of Things (IoT). *Procedia Computer Science,* 50: 99–102.
https://doi.org/10.1016/j.procs.2015.04.067

Bekara, S. 2014. Security Issues and Challenges for the IoT-Based Smart Grid. *Procedia Computer Science,* 34: 532–537.
https://doi.org/10.1016/j.procs.2014.07.064

Berman, J. J. 2013. Introduction. In *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information:* xix–xxvi. Boston: Morgan Kaufmann.

Borgohain, T., Kuman, U., & Sanyal, S. 2015. Survey of Security and Privacy Issues of Internet of Things. *International Journal of Advanced Networking and Applications,* 9(11): 20–26.

Boyd, D., & Crawford, K. 2012. Critical Questions for Big Data. *Information, Communication and Society,* 15(5): 662–679.
http://dx.doi.org/10.1080/1369118X.2012.678878

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A.V., & Rong, X. 2015. Data Mining for the Internet of Things: Literature Review and Challenges. *International Journal of Distributed Sensor Networks,* 11(8).
https://doi.org/10.1155/2015/431047

Chen, M., Mao, S., Zhang, Y., & Leung, V. 2014. *Big Data: Related Technologies, Challenges, and Future Prospects.* Cham, Switzerland: Springer International Publishing.
http://dx.doi.org/10.1007/978-3-319-06245-7

Chen, X-W., & Lin, X. 2014. Big Data Deep Learning: Challenges and Perspectives. *IEEE Access,* 2: 514–525.
http://dx.doi.org/10.1109/ACCESS.2014.2325029

Cox, M., & Ellsworth, D. 1997. *Managing Big Data for Scientific Visualization.* ACM SIGGRAPH '97, August 1997.

Finkle, J., & Volz, D. 2015. Database of 191 Million U.S. Voters Exposed on Internet: Researcher. *Reuters,* December 29, 2015. Accessed April 10, 2017:
http://uk.reuters.com/article/us-usa-voters-breach-idUKKBN0UB1E020151229

Gol, H. S. 2016. Integration of Wireless Sensor Network (WSN) and Internet of Things (IOT): Investigation of Its Security Challenges and Risks. *International Journal of Advanced Research in Computer Science and Software Engineering,* 6(1): 37–40.

Guo, B., Zhang, D., Wang, Z., Yu, Z., & Zhou, X. 2012. Opportunistic IoT: Exploring the Harmonious Interaction between Human and the Internet of Things. *Journal of Network and Computer Applications,* 36(6): 1531–1539.
https://doi.org/10.1016/j.jnca.2012.12.028

Han, G., Chan, S., Shu, L., & Hu, J. 2014. Security and Privacy in Internet of Things: Methods, Architectures, and Solutions. *Security and Communication Networks,* 7(11): 2181–2181.
http://dx.doi.org/10.1002/sec.1065

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. 2015. The Rise of "Big Data" on Cloud Computing: Review and Open Research Issues. *Information Systems,* 47: 98–115.
https://doi.org/10.1016/j.is.2014.07.006

Henze, M., Hermerschmidt, L., Kerpen, D., Häußling, R., Rumpe, B., & Wehrle, K. 2016. A Comprehensive Approach to Privacy in the Cloud-Based Internet of Things. *Future Generation Computer Systems,* 56: 701–718.
https://doi.org/10.1016/j.future.2015.09.016

Kelly, G. 2014. eBay Suffers Massive Security Breach, All Users Must Change their Passwords. *Forbes,* May 21, 2014. Accessed April 10, 2017:
http://www.forbes.com/sites/gordonkelly/2014/05/21/ebay-suffers-massive-security-breach-all-users-must-their-change-passwords/

Krotoszynski, R. J. Jr. 2015. Reconciling Privacy and Speech in the Era of Big Data: A Comparative Legal Analysis. *William and Mary Law Review,* 56(4): 1309.

Liu, C. 2014. External Integrity Verification for Outsourced Big Data in Cloud and IoT: A Big Picture. *Future Generation Computer Systems,* 49: 58–67.
http://dx.doi.org/10.1016/j.future.2014.08.007

Lu, X., Qu, Z., Li, Q., & Hui, P. 2015. Privacy Information Security Classification for Internet of Things Based on Internet Data. *International Journal of Distributed Sensor Networks,* 11(8).
http://dx.doi.org/10.1155/2015/932941

Malik, P. 2013. Governing Big Data: Principles and Practices. *IBM Journal of Research and Development,* 57(3/4): 1–13.
https://doi.org/10.1147/JRD.2013.2241359

Maras, M.-H. 2015. Internet of Things: Security and Privacy Implications. *International Data Privacy Law,* 5(2): 99–104.
https://doi.org/10.1093/idpl/ipv004

Mathews, A.W. 2015. Anthem: Hacked Database Included 78.8 Million People. *The Wall Street Journal,* February 24, 2015. Accessed April 10, 2017:
http://www.wsj.com/articles/anthem-hacked-database-included-78-8-million-people-1424807364

Matzner, T. 2014. Why Privacy Is Not Enough Privacy in the Context of "Ubiquitous Computing" and "Big Data". *Journal of Information, Communication & Ethics in Society,* 12(2): 93–106.
http://dx.doi.org/10.1108/JICES-08-2013-0030

Mills, E. 2011. SF Utilities Agency Warns of Potential Breach. *CNET,* June 2, 2011. Accessed April 10, 2017:
http://www.cnet.com/news/sf-utilities-agency-warns-of-potential-breach/

Big Data and Individual Privacy in the Age of the Internet of Things

Mackenzie Adams

Nakashima, E. 2015. Chinese Breach Data of 4 Million Federal Workers. *The Washington Post,* June 4, 2015. Accessed April 10, 2017:
https://www.washingtonpost.com/world/national-security/chinese-hackers-breach-federal-governments-personnel-office/2015/06/04/889c0e52-0af7-11e5-95fd-d580f1c5d44e_story.html

Narayanan, A., & Shmatikov, V. 2008. Robust De-Anonymization of Large Sparse Datasets. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy,* SP '08: 111–125.

O'Leary, D. 2013. Artificial Intelligence and Big Data. *IEEE Intelligent Systems,* 28(2): 96-99.
https://doi.org/10.1109/MIS.2013.39

Padgeta, J., & Vasconcelosb, W. W. 2015. Policy-Carrying Data: A Step Towards Transparent Data Sharing. *Procedia Computer Science,* 52: 59–66.
https://doi.org/10.1016/j.procs.2015.05.020

Perera, C., Ranjan, R., Wang, L., Khan, S. U., & Zomaya, A. Y. 2015. Big Data Privacy in the Internet of Things Era. *IT Professional,* 17(3): 32–39.
https://doi.org/10.1109/MITP.2015.34

Ponemon Institute. 2014. *The Cost of Data Breach Study.* Traverse City, MI: Ponemon Institute.

Punagin, S., & Arya, A. 2015. Privacy in the Age of Pervasive Internet and Big Data Analytics: Challenges and Opportunities. *International Journal of Modern Education and Computer Science,* 7(7): 36–47.
http://dx.doi.org/10.5815/ijmecs.2015.07.05

Rghioui, A., Aziza, L., Elouaai, F., & Bouhorma, M. 2015. Protecting E-Healthcare Data Privacy for Internet of Things Based Wireless Body Area Network. *Research Journal of Applied Sciences, Engineering and Technology,* 9(10): 876–885.

Russo, G., Marsigalia, B., Evangelista, F., Palmaccio, M., & Maggioni, M. 2015. Exploring Regulations and Scope of the Internet of Things in Contemporary Companies: A First Literature Analysis. *Journal of Innovation and Entrepreneurship,* 4(11).
http://dx.doi.org/10.1186/s13731-015-0025-5

Samani, A., Ghenniwa, H. H., & Wahaishi, A. 2015. Privacy in Internet of Things: A Model and Protection Framework. *Procedia Computer Science,* 52: 606–613.
https://doi.org/10.1016/j.procs.2015.05.046

Saroiu, S., Wolman, A., & Agarwal, S. 2015. Policy-Carrying Data: A Privacy Abstraction for Attaching Terms of Service to Mobile Data. In *HotMobile '15: Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications:* 129–134. New York: Association for Computing Machinery.
https://doi.org/10.1145/2699343.2699357

Schroeck, M., Shockley, R., Smart, D., Romero-Morales, J., & Tufano, P. 2012. *Analytics: The Real-World Use of Big Data.* IBM Global Business Services.

Shukla, S. 2015. Editorial: Big Data, Internet of Things, Cybersecurity: A New Trinity of Embedded Systems Research. *ACM Transactions on Embedded Computing Systems,* 14(4): 1–2.
https://doi.org/10.1145/2820608

Sen, R., & Borle, S. 2015. Estimating the Contextual Risk of Data Breach: An Empirical Approach. *Journal of Management Information Systems,* 32(2): 314–341.
http://dx.doi.org/10.1080/07421222.2015.1063315

Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porsini, A. 2015. Security, Privacy and Trust in Internet of Things: The Road Ahead. *Computer Networks,* 76: 146–164.
https://doi.org/10.1016/j.comnet.2014.11.008

Silver-Greenberg, J., Goldstein, M., & Perlroth, N. 2014. JPMorgan Chase Hacking Affects 76 Million Households. *The New York Times,* October 2, 2014. Accessed April 10, 2017:
http://dealbook.nytimes.com/2014/10/02/jpmorgan-discovers-further-cyber-security-issues/

Singel, R. 2009. Probe Targets Archives' Handling of Data on 70 Million Vets. *Wired,* October 1, 2009. Accessed April 10, 2017:
http://www.wired.com/2009/10/probe-targets-archives-handling-of-data-on-70-million-vets/

Smith, J., & Lee, M. 2015. Massive Hack of 70 Million Prisoner Phone Calls Indicates Violations of Attorney-Client Privilege. *The Intercept,* November 11, 2015. Accessed April 10, 2017:
https://theintercept.com/2015/11/11/securus-hack-prison-phone-company-exposes-thousands-of-calls-lawyers-and-clients/

Solomon, H. 2015. Popular Canadian-Owned Dating Sites Including Ashley Madison Hacked. *IT World Canada,* July 20, 2015. Accessed April 10, 2017:
http://www.itworldcanada.com/post/popular-canadian-dating-sites-including-ashley-maddison-hacked

Spiekermann, S., & Cranor, L. F. 2009. Engineering Privacy. *IEEE Transactions on Software Engineering,* 35(1): 67–82.
https://doi.org/10.1109/TSE.2008.88

Suciu, G., Suciu, V., Martian, A., Craciunescu, R., Vulpe, A., Marcu, I., Halunga, S., & Fratu, O. 2015. Big Data, Internet of Things and Cloud Convergence – An Architecture for Secure e-Health Applications. *Journal of Medical Systems,* 39(11): 1–8.
https://doi.org/10.1007/s10916-015-0327-y

Sweeney, L. 2002. k-anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* 10(5): 557–570.
https://doi.org/10.1142/S0218488502001648

Tsai, C. W., Lai, C. F., & Vasilakos, A. V. 2014. Future Internet of Things: Open Issues and Challenges. *Wireless Networks,* 20(8): 2201–2217.
https://doi.org/10.1007/s11276-014-0731-0

van de Pas J., & van Bussel G. 2015. 'Privacy Lost - and Found?' The Information Value Chain as a Model to Meet Citizens' Concerns. *The Electronic Journal Information Systems Evaluation,* 18(2): 185–195.

Weber, R.H. 2015. Internet of Things: Privacy Issues Revisited. *Computer Law & Security Review,* 31(5): 618–627.
https://doi.org/10.1016/j.clsr.2015.07.002

Westin, A. F. 1968. Privacy and Freedom. *Washington and Lee Law Review,* 25(1): 166–170.

Ziegeldorf, J. H., Morchon, O. G., & Wehrle, K. 2014. Privacy in the Internet of Things: Threats and Challenges. *Security and Communication Networks,* 7(12): 2728–2742.
http://dx.doi.org/10.1002/sec.795

Citation: Adams, M. 2017. Big Data and Individual Privacy in the Age of the Internet of Things. *Technology Innovation Management Review,* 7(4): 12–24.
http://timreview.ca/article/1067

Keywords: cybersecurity, Internet of Things, IoT, big data, privacy, smart devices, data breaches

Combining Exploratory Analysis and Automated Analysis for Anomaly Detection in Real-Time Data Streams

Ahmed Shah, Ibrahim Abualhaol, Mahmoud Gad, and Michael Weiss

> *“Besides black art, there is only automation and mechanization.”*
>
> Federico Garcia Lorca (1898–1936)
> Poet, playwright, and theatre director

Security analysts can become overwhelmed with monitoring real-time security information that is important to help them defend their network. They also tend to focus on a limited portion of the alerts, and therefore risk missing important events and links between them. At the heart of the problem is the system that analysts use to detect, explore, and respond to cyber-attacks. Developers of security analysis systems face the challenge of developing a system that can present different sources of information at multiple levels of abstraction, while also creating a system that is intuitive to use. In this article, we examine the complementary nature of exploratory analysis and automated analysis by testing the development of a system that monitors real-time Border Gateway Protocol (BGP) traffic for anomalies that might indicate security threats. BGP is an essential component for supporting the infrastructure of the Internet; however, it is also highly vulnerable and can be hijacked by attackers to propagate spam or launch denial-of-service attacks. Some of the attack scenarios on the BGP infrastructure can be quite elaborate, and it is difficult, if not impossible, to fully automate the detection of such attacks. This article makes two contributions: i) it describes a prototype platform for computing indicators and threat alerts in real time and for visualizing the context of an alert, and ii) it discusses the interaction of exploratory analysis (visualization) and automated analysis. This article is relevant to students, security researchers, and developers who are interested in the development or use of real-time security monitoring systems. They will gain insights into the complementary aspects of automated analysis and exploratory analysis through the development of a real-time streaming system.

Introduction

Security analysts can easily become overwhelmed with information, which can lead them to neglect critical alerts. This problem is exemplified in the 2013 Target data breach, which is one of the largest security breaches in history: it exposed 40 million credit card accounts and 70 million of the retailer’s customer profiles (Krebs, 2013). A forensic analysis of the attack (US Senate, 2014) found that the security monitoring systems put in place by Target had detected many of the key intrusion attempts during the attack; however, Target’s analysts were simply overwhelmed by the volume of alerts produced by the system and missed the early warning signs that a major attack was underway.

Indeed, analysts are “bombarded with alerts”, receiving so many that “they just don’t respond to everything” (Finkle & Heavey, 2014). Also, security analysts tend to focus on a limited portion of the alerts and therefore risk missing important events and relationships (Pierazzi et al., 2016). At the heart of the problem is the system that these analysts use to detect, explore, and respond to unanticipated and anticipated cyber-attacks. Generally speaking, developers of a security analysis platform (such as an intrusion detection system [IDS] or a security information and event management [SIEM] system) can face many challenges. Among them, a key challenge is how much data (e.g., raw traffic data) to present to analysts and to what extent the detection of anomalies should be automated by encoding detection rules into the system.

Combining Exploratory Analysis and Automated Analysis for Anomaly Detection

Ahmed Shah, Ibrahim Abualhaol, Mahmoud Gad, and Michael Weiss

The goal of this article is to examine the complementary nature of exploratory analysis and automated analysis for anomaly detection. For this purpose, we constructed a working prototype of a system to monitor real-time Border Gateway Protocol (BGP) traffic for security threats that combines both aspects. However, the application to BGP, as such, is not at the core of the present work: we simply used it to ground our work in a real-world context. The construction of the prototype produced two outcomes: i) categorization of attacks and indicators related to BGP derived from known threat scenarios and selection of indicators used in the prototype, and ii) an operationalization of the indicators and alerts (automation) and their visualization (exploration).

The findings presented in this article are most relevant to developers of systems for security monitoring. They face the challenge of developing intuitive systems for security analysts who are presented with different sources of information at multiple levels of abstraction (Corona et al., 2009). Developers also need to present this information at a human level of understanding that enables analysts to take appropriate and timely action (Corona et al., 2009). When analysts succeed in detecting “weak signals” (Fink et al., 2005) and acting on them early, their ability to manage security risks is greatly enlarged. It allows them to anticipate future attacks, rather than just reacting as they are detected.

This article is organized into four sections. We first review the literature on modes of analysis, the Border Gateway Protocol (BGP), and indicators and detection techniques for BGP attacks. We then describe the creation of a prototype system that combines exploratory analysis and automated analysis. Next, we examine the trade-off between exploratory analysis and automated analysis. We conclude by discussing lessons from the research that can be applied to the development of real-time security monitoring systems.

Literature Review

Modes of analysis

The automated analysis of network traffic works well for relatively stable environments. However, modern networks are growing in complexity and variability due to their dynamic and heterogeneous nature. This environment can create unstable systems in which the rules used by automated analysis become obsolete over time. Independently of our work, Pierazzi and colleagues (2016) found that a hybrid approach of exploratory analysis and automated analysis is necessary for effective anomaly detection.

Visualizing the observed data can help validate the outcomes of automated analysis. A visual representation of the context of an attack enables verification (Is the automated analysis correct?) and validation (Is the automated analysis meaningful?). Visualization techniques allow people to see and comprehend large amounts of complex data (Riad et al., 2011). Visualization can be used for the iterative improvement of automation rules. It also helps with the further exploration of an alert by an analyst to see what aspects of detection can be automated.

Border Gateway Protocol

Management of worldwide Internet traffic is administered by tens of thousands of independent routing domain systems called autonomous systems (AS) (Biersack et al., 2012). An AS can be owned by network operators such as Internet Service Providers (ISPs). The Border Gateway Protocol (BGP) is an inter-domain routing protocol used for managing network reachability information between more than one AS (Rekhter et al., 2006). Although BGP can be thought of as the protocol “that makes the internet work” (Pepelnjak, 2007), it is also considered as “the Internet’s biggest security hole” (Zetter, 2008). Malicious actors have the potential to influence BGP to deny service, sniff communications, reroute traffic to malicious networks, and create network instabilities (Meinel, 2008). Abnormal routing behaviour can disrupt global or local bound Internet connectivity and stability (Li et al., 2014; Murphy, 2006).

Indicators and detection techniques

In a survey of anomaly detection techniques for BGP data, Al-Musawi (2015) identified key indicators that can be used to detect BGP attacks. Among the most common indicators were the “number of BGP updates” and “AS path length”. The most common analytical approaches were time series analysis, machine learning, and statistical pattern recognition including support vector machines, hidden Markov models, and naive Bayes models. Biersack and colleagues (2012) surveyed various visual analytics tools for BGP, including node-link diagrams, rank-charge graphs, timelines, matrices, maps, and charts.

Creating a Platform that Combines Exploration and Automation

In this section, we describe the outcomes obtained from constructing a prototype of the analysis platform: i) the categorization of attacks and indicators related to BGP, as derived from known threat scenarios and the selection of indicators used in the prototype, and ii) the

Combining Exploratory Analysis and Automated Analysis for Anomaly Detection

Ahmed Shah, Ibrahim Abualhaol, Mahmoud Gad, and Michael Weiss

operationalization of indicators and alerts (automation) and their visualization (exploration). The platform was created strictly using open source technologies such as Apache Spark for real-time stream processing, D3.js and Crossfilter.js for visualization, MongoDB for data storage, Kafka for internal message communication, Flask for creating an external API, and libBGPstream for BGP data stream extraction.

Categorizing attacks and indicators

To conduct any kind of security analytics, we need to identify the known types of attacks and their key indicators. One proven way to compile this information is to examine attack cases and extract common attack characteristics and indicators. For this work, we surveyed BGP attack cases that were described in published studies, some of which also included unintentional attacks (e.g., a misconfiguration), as exemplified in Box 1. Preference was given to cases that included a detailed forensic analysis that examined indicators could be used for anomaly detection. Cases were also given priority for in-depth study if the attack dataset was publicly accessible. Out of the 15 cases that were collected, five general scenarios were identified where BGP was used for attacks: distribution of spam, influence of worms, traffic redirection for theft, eavesdropping, and denial-of-service attacks.

Box 1. Example scenario of a BGP IP prefix hijack

One widely cited BGP disruption scenario is the IP prefix hijack of YouTube in 2008. This hijack resulted from a foreign telecommunications company misconfiguring their systems: Pakistan Telecom inadvertently prevented users from around the world from accessing YouTube for roughly two hours. Pakistan Telecom was attempting to restrict its users from accessing YouTube. However, they accidentally sent new routing information via BGP to PCCW – an ISP in Hong Kong – which then propagated the false routing information across the whole Internet. This propagation amounted to a denial-of-service (DoS) attack on YouTube. In a DoS attack, users might not be able to obtain access to the Internet or specific websites. This type of attack is also known as a prefix hijacking attack: the Pakistan Telecom AS "hijacked" all traffic destined to YouTube, which amounted to sending Internet traffic meant for YouTube to Pakistan Telecom instead. This scenario involved two types of indicators: a spike in the number of a number of routes that contain the Pakistan Telecom AS and a spike in the AS advertisements made by Pakistan Telecom. A detailed forensic analysis of the attack was published by RIPE (2008).

After reviewing many indicators in the literature, we identified three that were common to most scenarios and should be observed by any analyst interested in BGP attacks:

1. Number of AS announcements: a sharp increase in the number of announcements is typically a strong indicator of hijacking (irrespective of whether it is malicious or not).

2. AS path length: the length of AS paths (list of systems that a BGP route follows from a given AS to the AS that owns a given prefix). During attacks AS path length increases. An analyst can observe the baseline behaviour to determine a typical AS path length and then use it to set a threshold, above which an alert should be thrown.

3. Multiple-origin AS (MOAS) conflict: more than one AS is claiming to be the owner of a given prefix. Any prefix should only be owned by one AS.

Operationalization of indicators and alerts (automation)

Figure 1 is a high-level flow diagram showing the main modules of the analysis platform that was constructed and the stages of information flow through the different modules. There are three main stages:

1. Input: collection of real-time data. In the BGP case study, BGP traffic is obtained from public data sources known as RIPE collectors, which archive BGP traffic data from around the world. The platform can either process BGP traffic obtained from collectors directly or use data replayed from an existing case file. The latter is important for training and validation purposes, as well as for forensic analysis of a particular attack.

2. Processing: extract, process, and dispatch features (i.e., key characteristics) of the data in real time (e.g., BGP announcements with information about the time of the announcement, the origin AS, and the AS path). The extracted features are sent to a message broker (Kafka), which will dispatch the information to different internal modules. MongoDB stores the features in a database, which will be used during

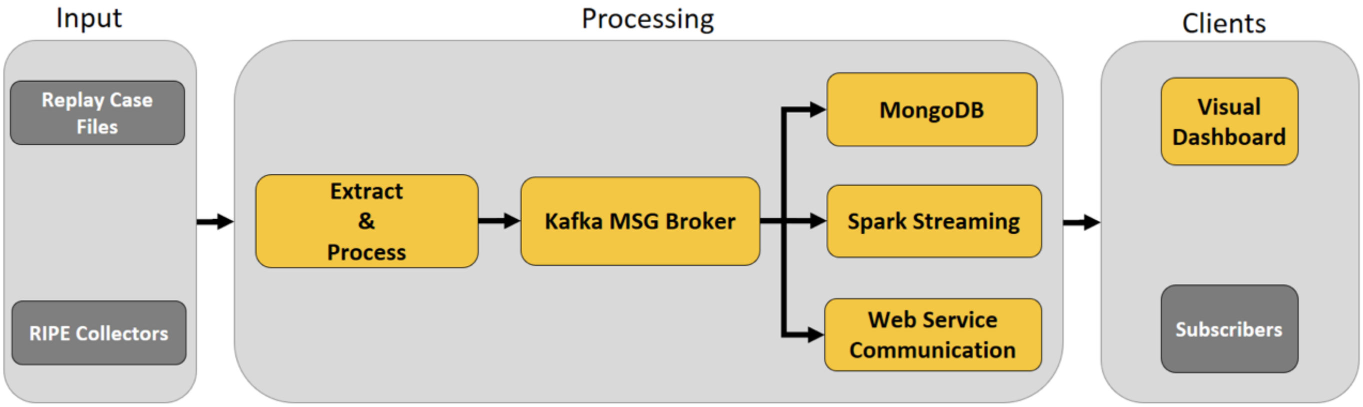


Figure 1. High-level architecture of the analysis platform

visualization. Apache Spark conducts further processing such as computing running averages of indicators or comparing indicators to thresholds. Web service communication provides an interface to external analytics systems.

3. Clients: clients include a visual dashboard – where alerts and indicators are visually presented to the user – or external systems that can subscribe to alerts.

Visualization (exploration)
Figure 2 shows the user interface (visual dashboard) of the analysis platform with the various visualization components. The visual dashboard contains three main sections:

A. Live Monitor: provides a simple status summary of real-time data stream ingestion.

B. Configure: a command-line-based interface for setting configuration parameters for controlling the input data stream (e.g., which IP prefix to monitor) and setting indicator thresholds.

C. Drill Down: provides a visual interactive dashboard on the data being ingested. This includes displaying recent alerts and providing interactive visualizations of the context of a given alert using timelines, histograms, and other graph types of the indicators that are being monitored.

Through the drill-down capability, the analyst can explore the context of a particular alert. They can zoom into a particular time range, showing only events and data related to that time interval, such as around a spike in a given indicator (e.g., the number of AS announcement). They can see when a given indicator is either unusually high or low by selecting the corresponding value or value range in a histogram component, upon which the other visualization components will be updated to show only corresponding values. For example, selecting just the high values for AS path length will reveal which AS and which prefixes were associated with long AS path lengths. Given that AS path lengths are generally short, a long AS path length may indicate a hijacking attack. By inspecting the origin AS of a long AS path, the analyst can quickly conclude which AS might be the source of the attack.

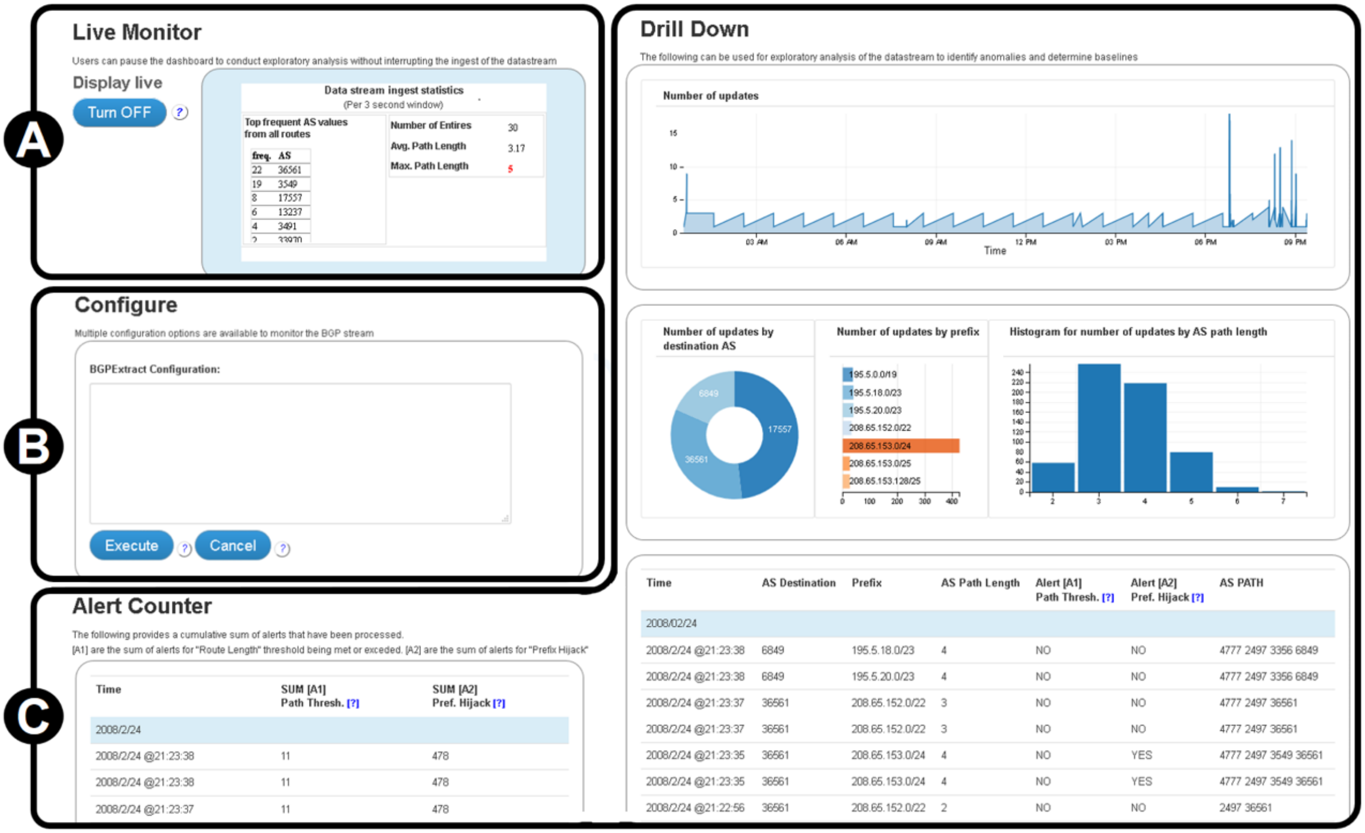
Figure 2 shows the results of the analysis platform replaying the YouTube 2008 IP Prefix hijacking case. The "number of updates" graph shows that there is a long period of time, from approximately 12:00am to 6:00pm, when updates are infrequent. For an analyst, this stable network activity could be considered a baseline that indicates that nothing beyond normal activity is occurring. When the IP prefix hijacking occurred (at approximately 6:30pm), there was a large increase in the frequency of updates, which may indicate an anomaly that the analyst should explore.

Trade-Offs between Exploratory Analysis and Automated Analysis

Figure 3 illustrates the interplay of automation and visualization. Automation (on the left) is about creating rules according to which real-time alerts will be raised. Alerts will be shown to an analyst in a dashboard. Visualization (on the right) is about providing the analyst with the ability to interactively explore the data associated with alerts (e.g., focus the analysis on specific time ranges or examine at which times a given indicator displayed unusually low or high values). The exploration of data might suggest patterns in the data (e.g., spikes

Combining Exploratory Analysis and Automated Analysis for Anomaly Detection

Ahmed Shah, Ibrahim Abualhaol, Mahmoud Gad, and Michael Weiss



Live Monitor

Users can pause the dashboard to conduct exploratory analysis without interrupting the ingest of the datastream

Display live

Data stream ingest statistics
(Per 3 second window)

Top frequent AS values from all routes

freq.	AS
22	36561
19	3549
8	17557
6	13237
4	3491
2	33970

Number of Entires	30
Avg. Path Length	3.17
Max. Path Length	5

Configure

Multiple configuration options are available to monitor the BGP stream

BGPExtract Configuration:

Alert Counter

The following provides a cumulative sum of alerts that have been processed.
[A1] are the sum of alerts for "Route Length" threshold being met or exceded. [A2] are the sum of alerts for "Prefix Hijack"

Time	SUM [A1] Path Thresh. [?]	SUM [A2] Pref. Hijack [?]
2008/2/24		
2008/2/24 @21:23:38	11	478
2008/2/24 @21:23:38	11	478
2008/2/24 @21:23:37	11	478

Drill Down

The following can be used for exploratory analysis of the datastream to identify anomalies and determine baselines

Time	AS Destination	Prefix	AS Path Length	Alert [A1] Path Thresh. [?]	Alert [A2] Pref. Hijack [?]	AS PATH
2008/02/24						
2008/2/24 @21:23:38	6849	195.5.18.0/23	4	NO	NO	4777 2497 3356 6849
2008/2/24 @21:23:38	6849	195.5.20.0/23	4	NO	NO	4777 2497 3356 6849
2008/2/24 @21:23:37	36561	208.65.152.0/22	3	NO	NO	4777 2497 36561
2008/2/24 @21:23:37	36561	208.65.152.0/22	3	NO	NO	4777 2497 36561
2008/2/24 @21:23:35	36561	208.65.153.0/24	4	NO	YES	4777 2497 3549 36561
2008/2/24 @21:23:35	36561	208.65.153.0/24	4	NO	YES	4777 2497 3549 36561
2008/2/24 @21:22:56	36561	208.65.152.0/22	2	NO	NO	2497 36561

Figure 2. User interface of the prototype analysis platform

in a given indicator) that may indicate potential attacks and should be codified into rules (e.g., new or revised thresholds attached to an indicator). This exploration helps build confidence in both the correctness and meaningfulness of the alerts.

On the one hand, we want the alerts and indicator thresholds to be correct. Only then analysts can be expected to rely on them. For example, when showing the AS involved in a prefix highjack attack, the designer of the dashboard may inadvertently be showing destination AS, rather than origin AS. In the case of a prefix highjack, however, only the origin AS will provide insights into which AS may be the source of the problem (such as the Pakistan Telecom AS in the YouTube scenario described in Box 1 and shown in Figure 2). A careful comparison of a known scenario against the values of the indicators in the dashboard can help detect such design errors.

On the other hand, we want the information provided to analysts to be meaningful. For example, if the threshold for an alert is set too low, too many alerts will be generated, overwhelming the analysts. Again, it may be difficult to determine the right threshold beforehand. However, by exploring the data, the analysts will be able to identify typical value ranges and thus suggest appropriate thresholds.

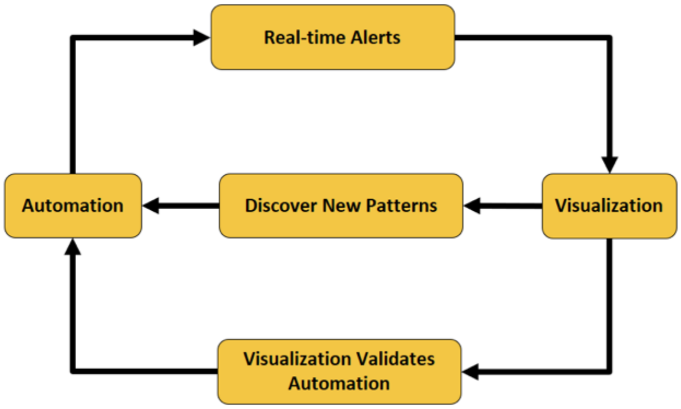


Figure 3. Interplay of automation and visualization

Combining Exploratory Analysis and Automated Analysis for Anomaly Detection

Ahmed Shah, Ibrahim Abualhaol, Mahmoud Gad, and Michael Weiss

During initial development of the analysis platform there was a focus on developing automation rules. When visualization components were added, relationships between indicators and anomalous behaviour in case studies became easier to identify, which then set the path for developing relevant automation rules. Automated approaches for anomaly detection should, therefore, be combined with preliminary explorations of the observed environment and data.

Conclusion

In this article, we explored the interplay between exploratory analysis and automated analysis. We described an experimental system for monitoring real-time data streams that combined exploratory analysis and automated analysis. The prototype incorporated both traditional rule-based mechanisms for detecting anomalies in data streams and interactive tools for discovering new anomalies and validating detection rules. Developers of real-time security monitoring systems can take the lessons from this research to reinforce the importance of how exploration and automation complement each other. Future work may include creating a real-time security information management system (SIEM) that uses machine learning to identify baseline patterns and potential attack patterns for processing data streams while also developing visualization components to tune algorithm accuracy.

About the Authors

Ahmed Shah holds a BEng in Software Engineering from Lakehead University in Thunder Bay, Canada, and a MEng in Technology Innovation Management from Carleton University in Ottawa, Canada. Ahmed has experience working in a wide variety of research roles at the VENUS Cybersecurity Corporation, the Global Cybersecurity Resource, and Carleton University.

Ibrahim Abualhaol is a Research Scientist at Larus Technologies and an Adjunct Professor at Carleton University in Ottawa, Canada. He holds a BSc, an MSc, and a PhD in Electrical and Computer Engineering. He is a senior member of IEEE and a Professional Engineer (P.Eng) in Ontario, Canada. His research interests include real-time big-data analytics and its application in cybersecurity and wireless communication systems.

Mahmoud M. Gad is a Research Scientist at the VENUS Cybersecurity Corporation. He holds a PhD in Electrical and Computer Engineering from the University of Ottawa in Canada. Additionally, he holds an MSc in ECE from the University of Maryland in College Park, United States. His research interests include big-data analytics for cybersecurity, cyber-physical system risk assessment, cybercrime markets, and analysis of large-scale networks.

Michael Weiss holds a faculty appointment in the Department of Systems and Computer Engineering at Carleton University in Ottawa, Canada, and he is a member of the Technology Innovation Management program. His research interests include open source, ecosystems, mashups, patterns, and social network analysis. Michael has published on the evolution of open source business, mashups, platforms, and technology entrepreneurship.

Combining Exploratory Analysis and Automated Analysis for Anomaly Detection

Ahmed Shah, Ibrahim Abualhaol, Mahmoud Gad, and Michael Weiss

References

Al-Musawi, B. 2015. *BGP Anomaly Detection Techniques: A Survey. Center for Advanced Internet Architectures (CAIA) Seminar.* Melbourne, Australia: Swinburne University of Technology.

Biersack, E., Jacquemart, Q., Fischer, F., Fuchs, J., Thonnard, O., Theodoridis, G., Tzovaras, D., & Vervier, P.-A. 2012. Visual Analytics for BGP Monitoring and Prefix Hijacking Identification. *IEEE Network Magazine,* 26(6): 33–39. https://doi.org/10.1109/MNET.2012.6375891

Corona, I., Giacinto, G., Mazzariello, C., Roli, F., & Sansone, C. 2009. Information Fusion for Computer Security: State of the Art and Open Issues. *Information Fusion,* 10(4): 274–284. http://doi.org/10.1016/j.inffus.2009.03.001

Fink, A., Marr, B., Siebe, A., & Kuhle, J. P. 2005. The Future Scorecard: Combining External and Internal Scenarios to Create Strategic Foresight. *Management Decision,* 43(3): 360–381. http://dx.doi.org/10.1108/00251740510589751

Finkle, J., & Heavey, S. 2014. Target Says It Declined to Act on Early Alert of Cyber Breach. Reuters, March 13, 2014. Accessed April 10, 2017: http://www.reuters.com/article/us-target-breach-idUSBREA2C14F20140313.

Krebs, B. 2013. Sources: Target Investigating Data Breach. *Krebs on Security,* December 18, 2013. Accessed April 10, 2017: https://krebsonsecurity.com/2013/12/sources-target-investigating-data-breach/

Li, Y., Xing, H.-J., Hua, Q., Wang, X.-Z., Batta, P., Haeri, S., and Trajkovic, L. 2014. *Classification of BGP Anomalies Using Decision Trees and Fuzzy Rough Sets.* Paper prseented at the IEEE International Conference on Systems, Man and Cybernetics (SMC), October 5–8, 2014, San Diego, CA. https://doi.org/10.1109/SMC.2014.6974096

Meinel, C. 2008. Attacking and Defending the Internet with Border Gateway Protocol (BGP). *Cisco Press,* August 25, 2008. Accessed April 10, 2017: http://www.ciscopress.com/articles/article.asp?p=1237179

Murphy, S. 2006. BGP Security Vulnerabilities Analysis. RFC 4272. *The Internet Society,* January 2006. Accessed April 10, 2017: https://tools.ietf.org/html/rfc4272

Pepelnjak, I. 2007. BGP Essentials: The Protocol that Makes the Internet Work. *SearchTelecom.* Accessed April 10, 2017: http://searchtelecom.techtarget.com/feature/BGP-essentials-The-protocol-that-makes-the-Internet-work

Pierazzi, D., Casolari, S., Colajanni, M., & Marchetti, M. 2016. Exploratory Security Analytics for Anomaly Detection. *Computers and Security,* 56(C): 28–49. https://doi.org/10.1016/j.cose.2015.10.003

Rekhter, Y., Li, T., & Hares, S. 2006. A Border Gateway Protocol 4 (BGP-4). RFC 4271. *The Internet Society,* January 2006. Accessed April 10, 2017: https://tools.ietf.org/html/rfc4271

Riad, A. E.-D., Elhenawy, I., Hassan, A., & Awadallah, N. 2011. Data Visualization Technique Framework for Intrusion Detection. *International Journal of Computer Science Issues,* 8(5): 440–443.

RIPE. 2008. YouTube Hijacking: A RIPE NCC RIS Case Study. *RIPE Network Coordination Centre,* March 17, 2008. Accessed April 10, 2017: https://www.ripe.net/publications/news/industry-developments/youtube-hijacking-a-ripe-ncc-ris-case-study

US Senate. 2014. *A "Kill Chain" Analysis of the 2013 Target Data Breach.* Washington, DC: United States Senate: Committee on Commerce, Science, and Transportation. http://rnc2.com/blog/wp-content/uploads/2014/11/Target%20Kill%20Chain%20Analysis.pdf

Zetter, K. 2008. Revealed: The Internet's Biggest Security Hole. *Wired,* August 26, 2008. Accessed April 10, 2017: https://www.wired.com/2008/08/revealed-the-in/

Citation: Shah, A., Abualhaol, I., Gad, M., & Weiss, M. 2017. Combining Exploratory Analysis and Automated Analysis for Anomaly Detection in Real-Time Data Streams. *Technology Innovation Management Review,* 7(4): 25–31. http://timreview.ca/article/1068

Keywords: cybersecurity, visualization, exploratory analysis, real-time data streams, anomaly detection

A Cybersecurity Risk Assessment Framework that Integrates Value-Sensitive Design

Aida Alvarenga and George Tanev

> *"The fact we have insecure embedded computers responsible for critical health functions should give pause to everyone involved. We hold banks responsible for security of a $10 online purchase, but we'll give medical device makers a free pass on not securing the devices responsible for our health or even our lives?"*
>
> Jay Radcliffe
> Cybersecurity researcher and diabetic
> who hacked his own insulin pump

Medical devices today are more effective and connected than ever before, saving more patient lives and making healthcare practitioner's jobs more efficient. But with this interconnectedness comes inherent concerns over increased cybersecurity vulnerabilities. Medical device cybersecurity has become an increasing concern for all relevant stakeholders including: patients, regulators, manufacturers, and healthcare practitioners. Although cybersecurity in medical devices has been covered in the literature, there is a gap in how to address cybersecurity concerns and assess risks in a way that brings value to all relevant stakeholders. In order to maximize the value created from cybersecurity risk mitigations, we review literature on the state of cybersecurity in the medical device industry, on cybersecurity risk management frameworks in the context of medical devices, and on how cybersecurity can be used as a value proposition. We then synthesize the key contributions of the literature into a framework that integrates cybersecurity value considerations for all relevant stakeholders into the risk mitigation process. This framework is subsequently applied to the hypothetical case of an insulin pump. Using this example case, we illustrate how medical device manufacturers can use the framework as a standardized method that can be applicable to medical devices at large. Our ultimate goal is to make cybersecurity risk mitigation an exploitable asset for manufacturers rather than a regulatory obligation.

Introduction

Advancements in technology have revolutionized the healthcare industry by making medical devices more productive, reducing the amount of human error, and enabling automation – all of which are helping healthcare practitioners treat more conditions and save more patient lives today than ever before (American Hospital Association, 2014). Connectivity of medical devices with the Internet and with other devices, however, has made them vulnerable to an array of cybersecurity threats (Burns et al., 2016). Since wireless interaction with these devices has become possible, they are no longer a standalone component in the clinical care process – they depend on connections and can interact with other devices remotely (Williams & Woodward, 2015). Over the next five years, interconnected health products are expected to be worth $285 billion in economic value – a number that is expected to grow exponentially over time (Harris, 2014). As medical devices become increasingly connected, and as some high-profile vulnerabilities are being exposed, cybersecurity of medical devices is garnering increased public, regulatory, and industry attention regarding cybersecurity risk and risk mitigation strategies. One dimension of these efforts that has not been readily addressed is how to convert these security efforts from an obligation to an asset that can maximize the value delivered to medical device stakeholders (MDPC, 2014). This value dimension of security requires a unique approach, first in the way that security risk is assessed and mitigated, and second in the way it affects stakeholders of medical devices themselves.

In this article, we first review the literature through the perspective of using security initiatives as a value proposition. We separated the literature into three streams: the current medical device cybersecurity landscape, medical device risk assessment, and cybersecurity as a value proposition. We then synthesized the results of the literature review into a framework that integrates stakeholder values with cybersecurity risk mitigation. This framework aims to provide a benchmark for medical device manufacturers when assessing cybersecurity concerns for a wide array of medical devices. In order to illustrate how the medical device cybersecurity risk assessment framework can be applied, and in particular how to choose risk controls that maximize value to key stakeholders, we applied it to the theoretical case of an insulin pump.

Literature Review

Medical devices: A unique cybersecurity landscape

A medical device is defined as "an instrument, apparatus, machine, implant, or similar article, including a component part or accessory... intended for the use in the diagnosis of disease or other conditions or in the cure, mitigation, treatment or prevention of disease" (Williams & Woodward, 2015). What makes medical devices unique is that security concerns involving these devices could directly affect treatments, safety, and even the life of a patient (Burns et al., 2016). For instance, implantable medical devices that have wireless connections – such as pacemakers, drug pumps, and defibrillators – if accessed, could leave control of the device in the hands of the hacker. Williams and Woodward (2015) identify key vulnerabilities faced by medical devices when it comes to cybersecurity. These include, but are not limited to: accessing the Internet through devices that are connected to internal networks, default admin passwords, web interfaces to infusion pumps, and web services that do not have encrypted communications.

Although no lives have been threatened yet through the hacking of a medical device, Jay Radcliffe, a cybersecurity researcher and diabetic proved that it was possible to hack and access his own insulin pump (Buntz, 2011). Even though attacks on medical devices with the goal of purposeful harm are expected to be very rare, the theoretical possibility cannot be ignored. Possible motivations for such attacks could be the acquisition of private information for financial gain, damage to the reputation of a manufacturer, or even terrorism (Maisel & Kohno, 2010). Attacks on healthcare IT networks have also become more prevalent in recent years. A SANS Institute (Filkins, 2014) report estimates that "up to 94% of medical organizations networks have been victims of a cyber-attack". This prevalence highlights the vulnerable environment that many medical devices are being exposed to. In light of this, the United States Federal Bureau of Investigation (FBI, 2015) has issued warnings that intrusions against medical devices and in the healthcare industry overall will increase due to lenient standards and the increased value of health data in the black market. Medical device manufacturers are also potential targets of cyber-attacks, and the "failure to properly prevent or patch cybersecurity risk may result in disapproval of a device, recall, or other regulatory or legal action" (Farrel & Hanet, 2016). Given these mounting cybersecurity concerns, the United States Food and Drug Administration (FDA) has issued a non-binding draft guidance for industry to follow in order to ensure the confidentiality, integrity, and availability of patient data (Maisel & Kohno, 2010). Some of the FDA's key recommendations include: identifying risks and vulnerabilities, determining risk levels and mitigation strategies, reporting vulnerabilities, and issuing routine updates or patches (FDA, 2016).

Security risk assessment for medical devices

The Medical Device Privacy Consortium (MDPC), which includes some of the largest medical device companies in the world, published a whitepaper proposing a security risk assessment framework for medical devices (MDPC, 2014). They identify a number of key issues to consider when applying existing security risk assessment frameworks to a medical device. For example, they found that existing methods focus primarily on patient safety risks (i.e., negative impacts to a patient's health), or that they assess impact too broadly. They also observed a lack of uniformity around security risk assessment across the medical device industry, and even within different business units. Due to these differences, the outcomes of these assessments are not always understood and create challenges when knowledge needs to be transferred between stakeholders. Furthermore, for medical devices, there is minimal experimental data on security risks and the probability of occurrence of harm, which creates challenges for producing accurate and consistent probability determinations MDPC (2014).

To resolve these issues, the security assessment framework proposed by the MDPC (2014) is based on four core ideas:

1. *Device focused:* Integrate common principles and language that are used in existing security standards in order to facilitate transferability and comprehension of information.

2. *All devices:* The framework is to be universally applicable to all medical devices, throughout the full product lifecycle.

3. *Tailored impact:* The framework will focus specifically on the impact to the confidentiality, integrity, and availability of information within the context of medical devices.

4. *Simplified probability:* Risk probability will be defined in a qualitative manner, focusing on the ability to exploit vulnerabilities associated with identified risk scenarios.

The MDPC framework requires manufacturers to identify threat sources and vulnerabilities, develop risk scenarios, assess exploitability, assess impact, obtain risk scores, and make decisions about how the risk can be mitigated. The framework provides a structured and straightforward approach to identifying security risks and scenarios that caters to the unique dimensions of the medical device industry. It provides the general goal of determining whether additional security controls are necessary to reduce the residual risk. The MDPC adapts the NIST 800-30 definition of security control for its application to medical devices as “The management, operational and technical controls (i.e., safeguards or countermeasures) prescribed for an information system and/or medical device to protect the confidentiality, integrity, and availability of the system and/or device and its information” MDPC (2014). The MDPC framework does not suggest a process or criteria for choosing the right security control for a given risk, given that the options are not singular, or trivial. One of the keys to success emphasized in the MDPC (2014) whitepaper is that manufacturers should strive to make product security an asset, not an obligation. This point highlights the need to integrate the value creation process into the security risk controls that are generated by the risk assessment process.

Wu and Eagles (2016) take the approach of leveraging medical device manufacturer’s proficiency with safety risk analysis (typically based on the ANSI/AAMI/ISO 14971 medical devices risk management standard) for cybersecurity risk analysis. They draw the parallel in the term “asset”, which is typically used indirectly in security standards, to the term “harm”, which is used in ANSI/AAMI/ISO 14971. Asset refers to the subject in need of protection, whereas harm implies that the subjects to be protected are people, property, or the environment. Wu and Eagles base the assessment process on a causal chain analogy which breaks down all of the stages and factors in an attack.

Wu and Eagles’ (2016) risk assessment approach takes a similar but significantly more detailed approach than the framework proposed by MDPC (2014) . Some of the key differences are their elaboration of risk control considerations, their emphasis on linking cybersecurity risk to safety risk, and their guidance on documentation. However, there are differences between safety and cybersecurity risks within the context of medical devices. Safety risks, as defined in the ISO 14971 (2010) standard, relate specifically to unintended hazards that can result in potential harm to patients. Cybersecurity risks relate specifically to intentional threats to the confidentiality, integrity, and availability of information of a medical device. Cybersecurity risks could therefore have safety impacts if they represent a source of harm to a patient. The security risk controls are not different from safety controls, given that they both aim to reduce the likelihood or severity of an event. As described in the MDPC (2014) framework, the process of choosing controls is not trivial, especially when there are multiple control options. Wu and Eagles (2016) highlight that cybersecurity controls need to be balanced against usability, which is also articulated in the FDA’s guidance (FDA, 2016). An example of the tradeoff is the use of a password to access information on a medical device, which could result in a delay of treatment. The impact of security on usability is important to consider, but Wu and Eagles, as well as the FDA, frame it as a tradeoff. This view overlooks the fact that security controls can be implemented in a way that adds value to stakeholders. This value could potentially be added in usability, by adding a fingerprint reader for both authentication and turning on the display, peace of mind, by securing patients’ private information by encryption, or in other ways based on the type of the device. Wu and Eagles also stress the importance of articulating cybersecurity controls implemented by a manufacturer in order to communicate the value of these controls within their organization and to external stakeholders and externally. This articulation of controls is a challenge for many medical device manufacturers when dealing with regulatory bodies, customers, and other stakeholders (Denning et al., 2014). Wu and Eagles propose that cybersecurity assessment information should be structured as an assurance case to facilitate the review process. An assurance case is a communication method that organizes information in a systematic and structured way to articulate evidence and critical thinking, and it is traditionally applied to safety assessments (FDA, 2014). Wu and Eagles (2016) provide a template of a cybersecurity assurance case and propose that this assurance case can be used to articulate cybersecurity assessment to outside stakeholders, specifically to regulators, which has also been recommended by the FDA for infusion pump manufacturers.

The qualitative measure of risk probability is one of the major contributions of the MDPC framework and could also strengthen and simplify the risk assessment of Wu and Eagles. Wu and Eagles do not clearly articulate how a risk is graded or scored in order to determine whether or not the risk warrants further controls. The MDPC highlights that this is an existing challenge, which is why they present their qualitative security risk probability measure. Wu and Eagles do stress the importance of security usability, value, and articulation, which is only briefly mentioned by the MDPC. Together, these two frameworks provide a comprehensive approach to medical device cybersecurity risk mitigation and the consideration of the value that is being created.

Cybersecurity as a value proposition

As reported above, the MDPC (2014) risk assessment whitepaper recommends that medical device manufacturers should view cybersecurity as an asset, rather than an obligation. Related to this view, Denning and colleagues (2014) have applied the principles of value-sensitive design to security system design, and Tanev and colleagues (2015) propose an ecosystem value blueprint approach to including cybersecurity as part of the manufacturer's value proposition.

We define value as something that resonates with and is perceived as useful to a relevant stakeholder (Anderson et al, 2006). Beyond a mere listing of benefits, value must resonate with the stakeholder. The approach to cybersecurity system design taken by Denning and colleagues (2014) is based on the idea that the most effective design is the one that brings the most value to all stakeholders. They apply principles of value-sensitive design to first identify all stakeholders to a medical device and second to identify value dams and flows. They apply this approach to the security and access control system of implantable cardiac devices. The authors argue that medical device value is typically discussed in terms of security, privacy, and convenience, with other dimensions being overlooked. These value dimensions include human values such as trust, physical welfare, autonomy, and human dignity. With a more holistic approach to all stakeholder values, manufacturers could potentially produce more secure devices that deliver greater value. Maximizing the value created by security controls that are produced from the risk assessment process warrants this type of holistic analysis of value.

Some of Denning's earlier work applies value-sensitive design to the security and access control system of implantable cardiac devices based on the patient's perception of value (Denning et al., 2010). In Denning and colleagues' follow-up work (2014), they approached value from the perspective of 24 healthcare providers whom they asked to identify which one of six security design concepts they favoured most based on their value-sensitive design approach. The ultimate goal was to identify which security and access system design concept created the most value for stakeholders. The value-sensitive methodology used by Denning and colleagues is separated in two parts. In the first part, they identified direct and indirect stakeholders (healthcare providers) to implantable medical devices. In the second part, they conducted a workshop, which included a metaphor-generating session for key terms associated with medical devices and security, a "critiques and concerns" session about the security of implantable cardiac devices, and a question-based evaluation highlighting the security controls that the participant liked or disliked and would or would not recommend.

The goal of this approach was to gain an in depth understanding of what aspects of the different security and access control systems generated value (value flows), and what aspects generated concern (value dams).

One of the key takeaways from the metaphor generation stage is that different stakeholders conceptualize security concepts differently when translated into lay terms. It is important for researchers to analyze these metaphors and to understand whether they are positive or negative when conceptualized into laymen terms. For example, a metaphor for a medical device could be positive (e.g., life saver) or negative (e.g., site of infection).

By combining the information generated by the question-based evaluation of the security and access control systems, and the critiques and concerns, the researchers found that the fail-open/safety wristband was best received. This was chosen as the hypothetical design choice from the six options.

Tanev and colleagues (2015) emphasize the importance of medical device manufacturers leveraging cybersecurity as a valuable differentiator. They propose a cybersecurity value blueprint approach that visually identifies all relevant stakeholders as part of an ecosystem and all associated security vulnerabilities. These vulnerabilities could be manifested by the stakeholders themselves or could simply involve the stakeholder in the security risk scenario. In any case, once manufacturers identify all high-risk vulnerabilities, they develop a plan in collaboration with stakeholders to mitigate these risks. The value dimensions of these cybersecurity mitigation efforts are articulated through a visual blueprint of all stakeholders in the medical device ecosystem.

Proposed Framework for Cybersecurity Value Creation through Risk Mitigation

By synthesizing key contributions from our review of the literature, we propose an approach to integrating cybersecurity value propositions into the risk assessment process. The work of Tanev and colleagues (2015) provided the overall structure to identify key stakeholders and to resolve high-risk vulnerabilities by addressing the security value dimensions. The MDPC (2014) security assessment framework provides an approach to identifying these high-risk vulnerabilities that is specific to the context of medical devices. We also found that the consideration of value created by security risk controls needs to be integrated into the risk assessment process. The value created can be related to usability, privacy, safety or other factors. The value-sensitive design approach for security by Denning and colleagues (2014) provides a methodology in considering stakeholder values when presented with a set of risk control options. Figure 1 shows how these various sources were synthesized into our proposed framework for cybersecurity value creation through risk mitigation.

Our framework divides the risk mitigation process into four stages:

A. *Identify stakeholders and their ecosystem relationships:* All key stakeholders to the medical device manufacturer are identified, along with how they relate to each other within the ecosystem. Stakeholders can be grouped in one of the stakeholder groups. For example, intermediaries would represent anyone between the manufacturer and end customer, such as regulators, insurance companies, or healthcare providers. The overall goal is to identify all relevant stakeholders that could either affect, or be affected by, cybersecurity risks.

B. *Identify security risks to be addressed:* The proposed approach for identifying key risks is the MDPC (2014) medical device security assessment framework, which proposes a qualitative method for calculating the probabilities of security risks.

C. *Identify all possible risk controls:* For each security risk that requires mitigation, a list of risk controls is to be developed in collaboration with subject matter experts and stakeholders, taking relevant security standards and regulations into consideration.

D. *Choose risk controls using value-sensitive design:* When risk controls that meet all security requirements have been identified for a specific risk, a value-sensitive design approach is used to choose the control that generates the most value (or reduces the least amount of value) for relevant stakeholders. This approach requires a workshop with a sample of all relevant stakeholders. This involves ranking all risk controls for a risk and choosing the one that ranks the highest.

The goal of this framework is to integrate stakeholder identification (Tanev et al., 2015) and value-sensitive design (Denning et al., 2014) to a security risk assessment designed specifically for medical devices (MDPC, 2014). With this framework, we aim to produce a reproducible process for stakeholders to effectively address cybersecurity concerns while maximizing stakeholder value.

Applying the Framework to a Hypothetical Case

In this section, we illustrate how the framework could be applied using the hypothetical example of an insulin pump. An insulin pump is a small, portable device that helps people with diabetes regulate their blood glucose levels by continuously monitoring and delivering insulin into the bloodstream as needed to maintain target levels. Some insulin pumps have Internet connectivity to enable features such as improved monitoring, remote monitoring and record keeping, and software updates.

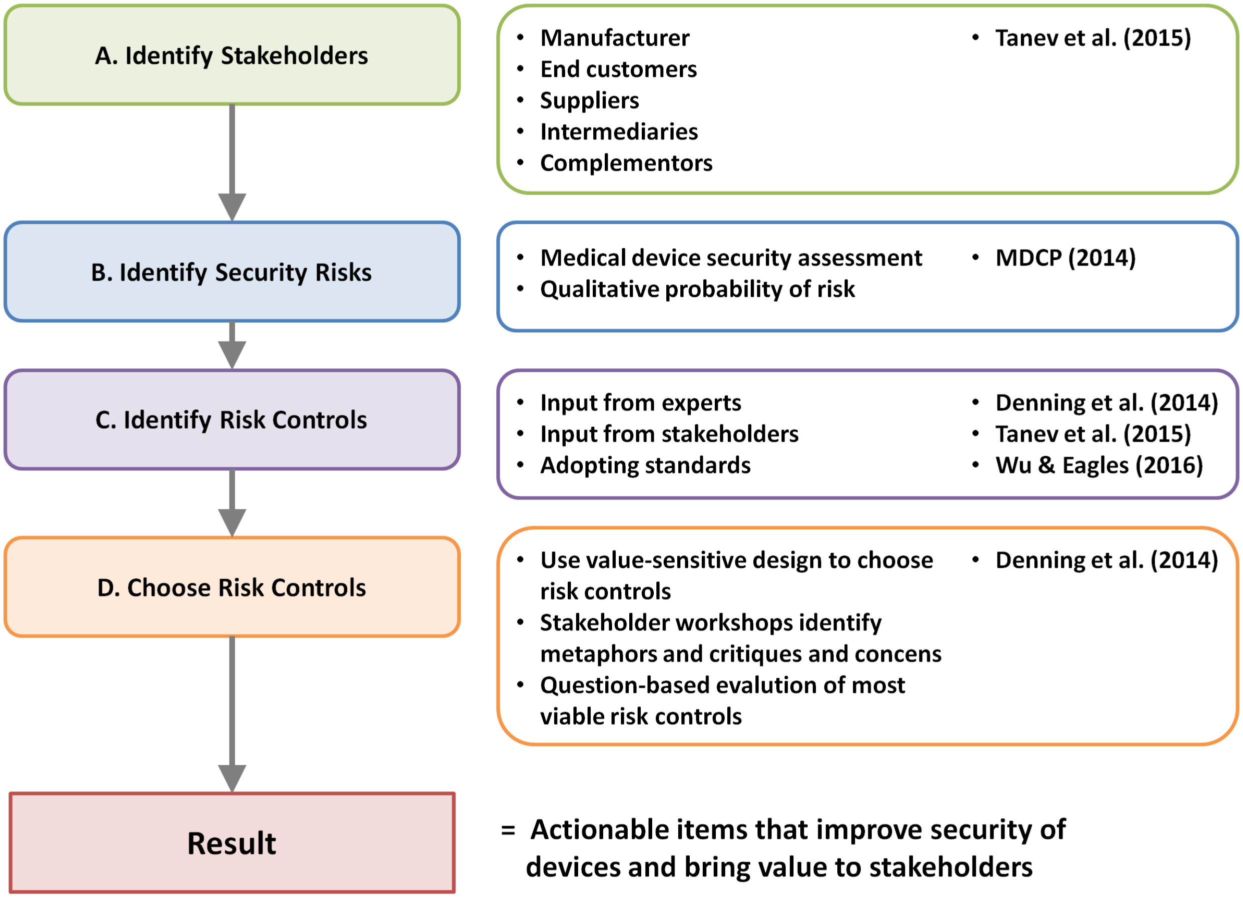

A. Identify Stakeholders
• Manufacturer
• End customers
• Suppliers
• Intermediaries
• Complementors
• Tanev et al. (2015)
B. Identify Security Risks
• Medical device security assessment
• Qualitative probability of risk
• MDCP (2014)
C. Identify Risk Controls
• Input from experts
• Input from stakeholders
• Adopting standards
• Denning et al. (2014)
• Tanev et al. (2015)
• Wu & Eagles (2016)
D. Choose Risk Controls
• Use value-sensitive design to choose risk controls
• Stakeholder workshops identify metaphors and critiques and concens
• Question-based evalution of most viable risk controls
• Denning et al. (2014)
Result
= Actionable items that improve security of devices and bring value to stakeholders

Figure 1. Framework for cybersecurity value creation through risk mitigation

During first three stages, the application of the framework to the insulin pump device gathers findings from the article by Paul, Kohno, and Klonoff (2011) who review key risks and possible controls for the specific case of an insulin pump. We used their article to derive existing knowledge from experts and incorporated it into the new framework. In practice, during the fourth stage, real workshops involving all relevant stakeholders would take place to identify stakeholder values and priorities. Here, given that our goal is simply to provide an example of how to approach this framework, we selected only a few key stakeholder groups and produced hypothetical data for stage four in order to illustrate the entire process.

Below, we organize the results of applying the framework to this case into subsections based the four stages of the framework, as outlined above and in Figure 1. We start by identifying stakeholders and their ecosystem relationships (Stage 1). We then identify the security risks that need to be addressed (Stage 2) and possible risk controls (Stage 3). Finally, we choose risk controls using the value-sensitive design approach (Stage 4).

A. Identify stakeholders and their ecosystem relationships
In the case of an insulin pump manufacturer, five key stakeholders were identified based on traditional stakeholders in a medical device ecosystem (Tanev et al., 2015):

1. *Manufacturers:* This group includes manufacturers of insulin pumps, or even different business units within the manufacturing organization. For example, the design team may have different goals than the engineers.

2. *Suppliers:* This group include both software and hardware suppliers.

3. *Complementors:* This group includes glucose monitor manufacturers, providers of insulin (the medicine used to treat diabetes), and database or cloud storage companies that work with the manufacturer.

4. *Intermediaries:* This group includes federal regulatory bodies that dictate the requirements and safety guidelines for devices as well as approve them for market release; Insurance companies that may fund the purchase of these devices for users; distributors of medical devices (e.g., hospitals or other agencies providing insulin pumps to patients); and healthcare providers (doctors and other practitioners who interact with the device but are not the end user).

5. *Users:* This group includes patients that have diabetes and use insulin pumps to regulate their glucose levels.

B. Identify security risks to be addressed

The insulin pump system under review (Paul et al., 2011) included a series of components: the insulin pump, a continuous glucose management system, a blood glucose monitor, and other devices (e.g., a mobile phone or computer). Two types of common security risks were chosen as examples given the type of insulin pump under review (Paul et al., 2011):

- **Risk 1:** Ensuring that remote control is only available to pre-approved individuals (i.e., the patient or their doctor) to maintain the integrity of system settings, to address system communication availability, and to ensure the software has not been altered without consent.

- **Risk 2:** Maintaining the integrity and confidentiality of data.

C. Identify all possible risk controls

Given the security risks, the manufacturer must decide what control to apply, if any. The following options for controlling the risks were identified:

Risk 1: Ensuring remote control is only accessed by pre-approved individuals

1. *Fail-safe physical interface:* Enables patient control when wireless communication fails (i.e., is lost, stolen, or interrupted).

2. *Wireless-enabling button:* Enables wireless communication on the device for short periods of time.

3. *Wireless-disabling switch:* Disables remote control, for example to start or stop insulin delivery when data is compromised or someone has interfered with the device.

Risk 2: Maintaining the integrity and confidentiality of data

1. *Encryption with button:* Along with encryption of data that follows the advanced encryption standard (Selent, 2010), a tactile button allows physicians to access the data in emergency situations.

2. *Encryption with infrared port:* Along with encryption of data that follows the advanced encryption standard (Selent, 2010), an infrared port interfaces with a data reader.

D. Choose risk controls using value-sensitive design with stakeholders

Following Denning and colleagues (2014), we identified stakeholders and simulated the steps suggested by the value-sensitive design process. The relevant stakeholders for this case study are: medical device manufacturers, patient's (end-users), and healthcare providers. Tables 1, 2, and 3 show the outcomes of metaphor generation and concern collection, question-based evaluation, and ranking and selection of risk controls for Risk 1. Below, we outline the steps followed in this stage for Risk 1 (Ensuring that remote control is only accessed by pre-approved individuals):

1. *Metaphor generation:* Ask stakeholders to generate metaphors for "insulin pumps" and "remote control access and security controls".

2. *Critiques and concerns:* Ask stakeholders to voice their concerns, fears, or insecurities about remote control of insulin pump technology.

3. *Question-based evaluation:* Ask stakeholders a series of questions (see Denning et al., 2014) about which concepts they like and dislike, which they would choose or recommend, etc.

4. *Rank and select risk controls:* Qualitatively analyze items 1 and 2 and quantitatively analyze item 3.

Table 1. Stakeholder metaphor generation and collection of concerns for Risk 1

Stakeholder	Fail-Safe Physical Interface	Wireless-Enabling Button	Wireless-Disabling Switch
Patient	• Might make device bigger and harder to wear • Pushing buttons accidentally	• Vulnerable if pushed by accident • Caregivers might not know that wireless needs to be enabled in emergencies	• Battery might be drained if turned on by accident • Caregivers might not know that wireless needs to be enabled in emergencies
Healthcare Provider	• Patients pushing buttons accidentally • Patients more likely to lose their programming module	• Not automatically clear to emergency caregivers that they need to press a button	• Not automatically clear to emergency caregivers that they need to press a button
Manufacturer	• More expensive to develop physical control interface	• Battery drains if patient accidentally presses button repeatedly or holds it down	• No concerns

Table 2. Stakeholder question-based evaluation for Risk 1

Stakeholder	Like	Dislike	Recommend	Do Not Recommend
Patient 1	• Fail-safe interface • Wireless-disabling switch	• None	• Fail-safe interface • Wireless-disabling switch	• None
Healthcare Provider	• Fail-safe interface • Wireless-enabling button • Wireless-disabling switch	• None	• Fail-safe interface • Wireless-disabling switch	• None
Manufacturer	• Wireless-enabling button • Wireless-disabling switch	• Fail-safe interface	• Wireless-enabling button • Wireless-disabling switch **Why:** Controlled access to programming of device.	• Fail-safe interface **Why:** Redundant feature that does not provide increased security against attacks.

Table 3. Ranking and selection of risk control for Risk 1. (Percentages are independent of each other.)

Risk Control	Like	Dislike	Recommend	Do Not Recommend
Fail-Safe Interface	66%	33%	66%	33%
Wireless-Enabling Button	66%	0%	33%	0%
Wireless-Disabling Switch	100%	0%	100%	0%

Table 4. Stakeholder metaphor generation and collection of concerns for Risk 2

Stakeholder	Encryption with Button	Encryption with Infrared Port
Patient	• No concerns	• Emergency staff may not have infrared device on hand • Someone with a stolen infrared device could read the data
Healthcare Provider	• No concerns	• Additional hardware necessary for emergency staff
Manufacturer	• Potential backdoor for decrypting information if tactile button is directly on the device • Encryption keys now need to be managed with partners	• Additional hardware required • Novel technology within this space • Encryption keys now need to be managed with partners

Table 5. Stakeholder question-based evaluation for Risk 2

Stakeholder	Like	Dislike	Recommend	Do Not Recommend
Patient 1	• Tactile button	• None	• Tactile button **Why:** Simple implementation and easy no hardware necessary.	• Infrared **Why:** Vulnerable if someone steals infrared reader
Healthcare Provider	• Tactile button	• Infrared port	• Tactile button **Why:** Doesn't require additional hardware for emergency staff	• Infrared port **Why:** Requires additional hardware for emergency staff
Manufacturer	• Tactile button	• Infrared port	• Tactile button **Why:** Simple and secure implementation both technologically and for emergency staff and caregivers	• Infrared port **Why:** Requires additional hardware for emergency staff, and additional resources to develop designated reader

Table 6. Ranking and selection of risk control for Risk 1. (Percentages are independent of each other.)

Risk Control	Like	Dislike	Recommend	Do Not Recommend
Encryption with Button	100%	0%	100%	0%
Encryption with Infrared Port	0%	66%	0%	100%

Tables 4, 5, and 6 show the outcomes of metaphor generation and concern collection, question-based evaluation, and ranking and selection of risk controls for Risk 2. Below, we outline the steps followed in this stage for Risk 2 (Maintaining the integrity and confidentiality of data):

1. *Metaphor generation:* Ask stakeholders to generate metaphors for "insulin pumps" and "patient glucose data".

2. *Critiques and concerns:* Ask stakeholders to voice their concerns, fears, or insecurities about the data integrity of glucose monitors and privacy of data in insulin pump technology.

3. *Question-based evaluation:* Ask stakeholders a series of questions (see Denning et al., 2014) about which concepts they like and dislike, which they would choose or recommend, etc.

4. *Rank and select risk controls:* Qualitatively analyze items 1 and 2 and quantitatively analyze item 3.

Discussion

In the hypothetical example, the results for Risk 1 (Ensuring remote control is only accessed by pre-approved individuals) show that the risk control that brought the most value to all three of the selected stakeholders was incorporating a switch to disable and enable wireless communication in the insulin pump. For Risk 2 (Maintaining the integrity and confidentiality of data), the risk control that was preferred by the three selected stakeholders was that of encrypting data with a tactile button instead of using an infrared port.

Our aim with this hypothetical application of the framework is to show how risk controls can be chosen in a way that considers the perceived value notion from a variety of stakeholders. In this illustrative example, we do not suggest that the stakeholders selected, the risks described, or the mitigation controls offered are best suited to making insulin pumps cybersecure. We acknowledge that there may be many more stakeholders, risks, and controls need to be accounted for when fully assessing insulin pumps and medical devices at large.

Our contribution is to showcase (at a small scale) how the proposed framework is applicable to a particular medical device. With this framework, we aim to make it easier to:

- Consider key stakeholders when evaluating and addressing cybersecurity risks in medical devices.

- Improve the safety of all stakeholders that are affected by these medical devices.

- Provide manufacturers with a framework that provides actionable items on how to improve their device's security in a way that brings value to their stakeholders (including themselves).

- Transform cybersecurity from a regulatory obligation into an asset (competitive advantage) for manufacturers.

- Evolve the medical device industry from its current position into one that puts cybersecurity at the forefront of its priorities.

Conclusion

In this article, we developed the key concepts necessary to articulate cybersecurity as a value proposition. Based on a review of the literature on the current landscape of medical device cybersecurity, on medical device risk mitigation, and on cybersecurity as a value proposition, we proposed a framework that integrates value articulation with the risk assessment and mitigation process. This framework takes into account the unique aspects of medical device security, the benefits of considering value creation when choosing risk controls, and the importance of perceiving value through the perspective of multiple stakeholders. The hypothetical case study of an insulin pump provided a practical example of applying the framework. It identified stakeholders, risks, potential mitigations, and the value that can be created for stakeholders for each mitigation. We used available resources to hypothetically analyze and choose risk mitigation options based on the perspectives of several key stakeholders. This framework is intended to be applied to any medical device with the purpose of articulating the value generated by cybersecurity within the context of medical device risk assessment.

About the Authors

Aida Alvarenga Castillo is a Master's student in the Technology Innovation Management program at Carleton University in Ottawa, Canada. Aida undertook her undergraduate studies at McGill University in Montreal, Canada, with a focus on Economics, Business Management, and Political Science. She has experience in the financial industry for well-established banks, in a business development role for a technology startup, and as an entrepreneur in launching her own family food business. Within the field of technology innovation, Aida's main interests are in financial technologies (FinTech) and innovation within the financial industry.

George Tanev is a Master's student in the Technology Innovation Management program at Carleton University in Ottawa, Canada. George holds a Master's of Science degree in Medicine and Technology from the Technical University of Denmark and a Bachelor of Engineering in Biomedical and Electrical Engineering from Carleton University. George has experience in the medical device industry and the air navigation services industry. His interests are in technology entrepreneurship, cybersecurity, medical device product development, signal processing, and data modelling.

References

American Hospital Association. 2014. A Message from the AHA: Considering Unique Cybersecurity Risks of Medical Devices Is Critical. *AHA News,* December 4, 2015. Accessed April 10, 2017: http://www.aha.org/advocacy-issues/141204cybersecurityrisksnews.shtml

Anderson, J., Narus, J., & van Rossum, W. 2006. Customer Value Propositions in Business Markets. *Harvard Business Review,* 84(3): 90–99.

Buntz, B. 2011. Insulin Pump Hacking: Sensationalism or Legitimate Threat? *Medical Device and Diagnostic Industry,* August 12, 2011. Accessed April 10, 2017: http://www.mddionline.com/blog/devicetalk/insulin-pump-hacking-sensationalism-or-legitimate-threat

Burns, A. J., Johnson, M. E. P., & Honeyman, P. 2016. A Brief Chronology of Medical Device Security. *Communications of the ACM,* 59(10): 66–72. http://dx.doi.org/10.1145/2890488

Denning, T., Borning, A., Friedman, B., Gill, B. T., Kohno, T., & Maisel, W. H. 2010. Patients, Pacemakers, and Implantable Defibrillators: Human Values and Security for Wireless Implantable Medical Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems:* 917–926. New York: Association for Computing Machinery.

Denning, T., Kramer, D. B., Friedman, B., Reynolds, M. R., Gill, B., & Kohno, T. 2014. CPS: Beyond Usability: Applying Value Sensitive Design Based Methods to Investigate Domain Characteristics for Security for Implantable Cardiac Devices. In *Proceedings of the 30th Annual Computer Security Applications Conference: ACSAC 2014:* 426–435. New York: Association for Computing Machinery. http://dx.doi.org/10.1145/2664243.2664289

Farrel, E., & Hanet, J. 2016. *Cybersecurity and Medical Devices: Electronic Medical Data Increases Product Liability Risk For Medical Device Manufacturers.* Toronto: Gowling WLG.

FBI. 2014. *FBI Cyber Division Bulletin: Health Care Systems and Medical Devices at Risk for Increased Cyber Intrusions.* Washington, DC: Federal Bureau of Investigation. https://publicintelligence.net/fbi-health-care-cyber-intrusions/

FDA. 2014. *FDA Case Study: An Infusion Pump Company Considers Risk Assessment and Mitigation.* Silver Spring, MD: U.S. Food and Drug Administration

FDA. 2016. *Draft Guidance: Postmarket Management of Cybersecurity in Medical Devices.* Silver Spring, MD: U.S. Food and Drug Administration

Harris, P. 2014. *The Prognosis for Healthcare Payers and Providers: Rising Cybersecurity Risks and Costs.* London: PricewaterhouseCoopers.

ISO. 2007. *ISO 14971: Medical Devices-Application of Risk Management to Medical Devices.* Geneva: International Organization for Standards.

Maisel, W. H., & Kohno, T. 2010. Improving the Security and Privacy of Implantable Medical Devices. *The New England Journal of Medicine,* 362(13): 1164–1166. http://dx.doi.org/10.1056/NEJMp1000745

MDPC. 2014. *Security Risk Assessment Framework for Medical Devices: A Medical Device Privacy Consortium White Paper.* Washington, DC: Medical Device Privacy Consortium.

Paul, N., Kohno, T., & Klonoff, D. C. 2011. A Review of the Security of Insulin Pump Infusion Systems. *Journal of Diabetes Science and Technology,* 5(6): 1557–1562.
https://doi.org/10.1177/193229681100500632

Filkins, B. 2014. *Health Care Cyberthreat Report: Widespread Compromises Detected, Compliance Nightmare on Horizon.* Bethesda, MD: SANS Institute.

Selent, D. 2010. Advanced Encryption Standard. *Rivier Academic Journal,* 6(2): 1–14.

Tanev, G., Tzolov, P., & Apiafi, R. 2015. A Value Blueprint Approach to Cybersecurity in Networked Medical Devices. *Technology Innovation Management Review,* 5(6): 17–25.
https://timreview.ca/article/903

Williams, P. A. H., & Woodward, A. J. 2015. Cybersecurity Vulnerabilities in Medical Devices: A Complex Environment and Multifaceted Problem. *Medical Devices: Evidence and Research,* 8: 305–316.
https://doi.org/10.2147/MDER.S50048

Wu, F., & Eagles, S. 2016. Cybersecurity for Medical Device Manufacturers: Ensuring Safety and Functionality. *Biomedical Instrumentation and Technology,* 50(1): 23–34.
http://dx.doi.org/10.2345/0899-8205-50.1.23

Citation: Alvarenga, A., & Tanev, G. 2017. A Cybersecurity Risk Assessment Framework that Integrates Value-Sensitive Design. *Technology Innovation Management Review*, 7(4): 32–43. http://timreview.ca/article/1069

Keywords: cybersecurity, risk assessment, framework, value-sensitive design, medical devices, value propositions

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

7. Include a 75-150 word biography.

8. List the references at the end of the article.

9. If there are any texts that would be of particular interest to readers, include their full title and URL in a "Recommended Reading" section.

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