Securing the Internet of Things: A Machine Learning Approach

Aziz Mohaisen
University of Central Florida
mohaisen@cs.ucf.edu

Joongheon Kim
Chung-Ang University
joongheon@cau.ac.kr

***Abstract*—The Internet of Things, IoT, is expected to revolutionize our lives, and their security of paramount importance. Machine learning, on the other hand, has found many applications in computer security in general, and IoT security in particular. In this tutorial, we review the state-of-the-art on machine learning applications for end-to-end Internet of Things systems security, by touching upon security issues at the hardware, software, and protocol level and how they are addressed using machine learning. Through this tutorial, we teach the various approaches used for performing behavior-based analyses used for engineering (or automatically extracting) features from the behavior of software associated with its use, the characteristics of hardware associated with baselines and behaviors of sensors, and the communication protocol-level artifacts. We review the existing literature on applications of such approach, including malware detection and behavior profiling and fingerprinting. We supplement our tutorial with a look into the applications of deep learning to such application area, and how advances in GPU-CPU co-design can further make those applications of machine learning real-time. Finally, we conclude with open directions, where the community has to come together and address the problems at hand.**

I. Executive Summary

Internet of things devices and their use are on the rise, and they are expected to revolutionize our way of living, especially in applications such as home networks. However, with this promise and potential come various security challenges, and among the most important of those challenges is the advanced persistent threats (APT), including malware. With APT, an adversary would not be able to only breach the security and privacy of users and home network devices, but as well affect their safety. Unlike mass-market malicious software, which can be easily identified using signatures of behavior, APTs are stealthy, requiring new capabilities that can better meet the needs in the home network with its specifics.

Today, even the simplest computer system found in IoT settings encompasses a complex convergence of software, hardware, and protocol designs and entities. The security and reliability of such entities is of paramount importance to those systems. For example, todays consumer-grade tablets found in smart home networks are featured with complex software, such as Android and its applications, sophisticated hardware featured with a large array of sensors and peripherals, and many interdependent communication protocols. The same features are associated with many Internet of Things devices, including home entertainment systems, smart TV, etc. The security and reliability of each and every of those entities in such systems is of paramount importance to their operation. To this end, to secure the Internet of Things, there are two general approaches: the clean slate approach and the incremental approach. The former approach is prohibitively expensive, and does not address the needs of legacy and well deployed systems, which are the advantages of the latter approach. Manually addressing security issues using the second approach does not scale to the size of the problems faced by todays complex computer systems, thus automation is required. One of the approaches to automated the security analysis and defense in computer systems is through machine learning: fully automated algorithms to detect security issues in the design and operation of computer systems, at the software, hardware and protocol level, and to guide defenses.

In this tutorial, we review the state-of-the-art on machine learning applications for end-to-end Internet of Things systems security, by touching upon security issues at the hardware, software, and protocol level and how they are addressed using machine learning. Through this tutorial, we will teach the various approaches used for performing behavior-based analyses used for engineering (or automatically extracting) features from the behavior of software associated with it use, the characteristics of hardware associated with baselines and behaviors of sensors, and the communication protocol-level artifacts. We survey the existing literature on applications of such approach, including malware detection and behavior profiling and fingerprinting. We supplement our tutorial with a look into the applications of deep learning to such application area, and how advances in GPU-CPU co-design can further make those applications of machine learning more practical. Finally, we conclude with open directions, where the community has to come together and address the problems at hand.

II. Outlines

Our tutorial is three parts: an introduction to the Internet of Things, Machine learning approach IoT security, and making machine learning practical (which includes open directions). Each of those parts will be over one hour (tentatively). The first part will be led by Dr. Kim, except for the "overview of the security threats landscape in the Internet of Things". The second part will be lead by Dr. Mohaisen, except the last avenue of application to hardware security. The final part will be co-led by both presenters; Dr. Kim will present the GPU computing section, and Dr. Mohaisen will present the rest.

An Introduction to the Internet of Things. In this 1-hour lecture we review basics of the Internet of Things, including the following topics: 1) A three layer-approach to IoT design: device, network and service, 2) An alternative approach to IoT research: software, hardware, and protocol, 3) An overview of the security threats landscape in the Internet of Things; a) Taxonomy of security threats (Attack surface analysis, Vulnerability/defense approach and limitations), b) Layer-specific security threats (Security threats due to the hardware, Security threats due to the hardware, Security threats due to protocols, A classical view of security threats: device, service), and 4) On the need for security automation.

Internet of Things Security: Machine Learning Approach. In this 1-hour lecture, we various aspects, including a motivation of why machine learning is the right approach to the problem at hand, what are the other alternatives that could be used for addressing the problem, and limitations of machine learning. Second, we cover common themes of machine learning applications and avenues to addressing the security of IoT, including behavioral analysis, features engineering, training and ground truth discovery, and evaluation and consensus. We supplement this general framework by analyzing three different applications, namely malware classification, protocol analysis and hardware security analysis.

Making Machine Learning Practical. In this final 1-hour lecture we cover practical considerations of machine learning for IoT. Namely, we consider practical considerations for multi-layered designs, the use of GPU to scale machine learning techniques, e.g., by boostrapping malware classification, and on the challenge of deploying machine learning in adversarial environments (i.e., adversarial machine learning). We sum up this lecture with open directions and concluding remarks.

III. Speakers

The presentation of this tutorial at IEEE ICC 2018 is the first time it is presented in a conference. We aim to present the tutorial at other conferences, with feedback collected at ICC 2018 to guide the future offerings of the tutorial. The authors have a rich experience on the problem at hand, including a large number of peer-reviewed publications, and existing funded projects from several industrial and government sponsors on systems security (including NVIDIA, Amazon, NSF, KRF, AFOSR, and AFRL). Sample publications on the topic include the work in [4], [3], [9], [2], [1], [5], [7], [10], [6], [8]

Joongheon Kim. Dr. Joongheon Kim has been an assistant professor of Computer Science and Engineering with Chung-Ang University, Seoul, Korea, since 2016. He received his B.S. and M.S. degrees from Korea University, Seoul, Korea, in 2004 and 2006, respectively, and his Ph.D. degree from the University of Southern California (USC), Los Angeles, CA, USA, in 2014. Before joining USC, he was a research engineer with LG Electronics, Seoul, Korea, from 2006 to 2009. He was also a systems engineer with Intel Corporation, Santa Clara, CA, USA, from 2013 to 2016. He has been published his research results in top-tier venues such as IEEE ICC, IEEE GLOBECOM, ACM MobiSys, ACM Multimedia, ACM HotPower, IEEE/ACM Transactions on Networking, IEEE Transactions on Broadcasting, IEEE Internet of Things Journal. He was awarded USC Annenberg Graduate Fellowship with his Ph.D. admission from USC, in 2009.

Aziz Mohaisen. Dr. Aziz Mohaisen earned his M.Sc. and Ph.D. degrees from the University of Minnesota in 2012. Currently, he is an Associate Professor of Computer Science at the University of Central Florida. Prior to joining Central Florida, he was an Assistant Professor at SUNY Buffalo (2015-2017), a Senior Research Scientist at Verisign Labs (2012-2015), and a Researcher at ETRI, a government-backed researched institute in South Korea (2007-2009). His research interests are broadly in cybersecurity, with applications to DDoS, malware, blockchain, and emerging networking technologies, such as Internet of Things. He was awarded the Summer Faculty Fellowship from the US AFOSR (2016), the Best Student Paper at ICDCS (2017), the Best Paper Award at WISA (2014), the Best Poster Award at IEEE CNS (2014), and a Doctoral Dissertation Fellowship from the University of Minnesota (2011). He was recognized for his service to IEEE INFOCOM (2017) and IEEE CNS (2016), and has been on the organizing committee of IEEE INFOCOM, IEEE ICDCS, IEEE CNS, IEEE PAC, SecureComm, ICCCN, HotWeb, MobiSys, AsiaCCS, etc. His research has been supported by various grants, and featured in MIT Technology Review, the New Scientist, Minnesota Daily, Slashdot, The Verge, Deep Dot Web, and Slate, etc. He is a member of ACM and a senior member of IEEE.

Acknowledgement. A. Mohaisen's work was partially supported by NSF grant CNS-1643207, NRF grant NRF-2016K1A1A2912757, NVIDIA GPU Grant Program, and AFRL Visiting Faculty Program (ICA2017-SVFRP-022). J. Kim was supported by KEPCO grant R17XA05-41 (2017). All opinions are of the authors and do not represent the sponsors.

References

[1] Fan Dang, Pengfei Zhou, Zhenhua Li, Ennan Zhai, Aziz Mohaisen, Qingfu Wen, , and Mo Li. Large-scale invisible attack on afc systems with nfc-equipped smartphones. In *In Proceedings of the 36th IEEE International Conf. on Computer Communications, INFOCOM 2017, Atlanta, GA, USA, May 1-4*, 2017.

[2] RhongHo Jang, Jeonil Kang, Aziz Mohaisen, and DaeHun Nyang. Rogue access point detector using characteristics of channel overlapping in 802.11n. In *37th IEEE International Conference on Distributed Computing Systems, ICDCS*, pages 2515–2520, 2017.

[3] Joongheon Kim and Aziz Mohaisen. Distributed and reliable decision-making for cloud-enabled mobile service platforms. *IJDSN*, 13(8), 2017.

[4] Hesham Mekky, Aziz Mohaisen, and Zhi-Li Zhang. Separation of benign and malicious network events for accurate malware family classification. In *Proc. of IEEE CNS*, 2015.

[5] Young Jong Mo, Joongheon Kim, Jong-Kook Kim, Aziz Mohaisen, and Woojoo Lee. Performance of deep learning computation with tensorflow software library in gpu-capable multicore computing platforms. In *The 9th International Conference on Ubiquitous and Future Networks*, 2017.

[6] Aziz Mohaisen. Towards automatic and lightweight detection and classification of malicious web contents. In *Third IEEE Workshop on Hot Topics in Web Systems and Technologies, HotWeb 2015, Washington, DC, USA, November 12-13, 2015*, pages 67–72, 2015.

[7] Aziz Mohaisen, Charles Kamhoua, Kevin Kwiat, and Laurent Njilla. Thwarting advanced persistent threats in internet of things networks. In *Technical Report, Air Force Research Lab*, 2017.

[8] DaeHun Nyang, Aziz Mohaisen, and Jeonil Kang. Keylogging-resistant visual authentication protocols. *IEEE Trans. Mob. Comput.*, 13(11):2566–2579, 2014.

[9] Feng Shen, Justin Del Vecchio, Aziz Mohaisen, Steven Y. Ko, and Lukasz Ziarek. Android malware detection using complex-flows. In *37th IEEE International Conference on Distributed Computing Systems, ICDCS*, pages 2430–2437, 2017.

[10] Jeffrey Spaulding, Jeman Park, Joongheon Kim, and Aziz Mohaisen. Proactive detection of algorithmically generated malicious domains. In *The 32nd International Conference on Information Networking, ICOIN 2018, Chiang Mai, Thailand*, 2018.