Data-Driven Intelligence for Characterizing Internet-scale IoT Exploitations

Nataliia Neshenko*, Martin Husák†*,Elias Bou-Harb*, Pavel Čeleda†, Sameera Al-Mulla ‡,
Claude Fachkha ‡
*Cyber Threat Intelligence Lab, Florida Atlantic University, USA
Email: {nneshenko2016, ebouharb}@fau.edu
†Institute of Computer Science, Masaryk University, Czech Republic
Email: {husakm, celeda}@ics.muni.cz
‡University of Dubai
Email: {salmulla, cfachkha}@ud.ac.ae

***Abstract*—While the security issue associated with the Internet-of-Things (IoT) continues to attract significant attention from the research and operational communities, the visibility of IoT security-related data hinders the prompt inference and remediation of IoT maliciousness. In an effort to address the IoT security problem at large, in this work, we extend passive monitoring and measurements by investigating network telescope data to infer and analyze malicious activities generated by compromised IoT devices deployed in various domains. Explicitly, we develop a data-driven approach to pinpoint exploited IoT devices, investigate and differentiate their illicit actions, and examine their hosting environments. More importantly, we conduct discussions with various entities to obtain IP allocation information, which further allows us to attribute IoT exploitations per business sector (i.e., education, financial, manufacturing, etc.). Our analysis draws upon 1.2 TB of darknet data that was collected from a /8 network telescope for a 1 day period. The outcome signifies an alarming number of compromised IoT devices. Notably, around 940 of them fell victims of DDoS attacks, while 55,000 IoT nodes were shown to be compromised, aggressively probing Internet-wide hosts. Additionally, we inferred alarming IoT exploitations in various critical sectors such as the manufacturing, financial and healthcare realms.**

I. Introduction

The Internet-of-Things (IoT) notion has been a buzz word, but will continue to represent an integral part of our contemporary life. From object recognition devices which promise to revolutionize physical therapy [1] to connected vehicles aiming at preventing the driver from deviating from proper trajectory paths or bumping into objects. IoT indeed possesses a significant impact on natural resources' integrity and consumption, along with monitoring environmental pollution and chemical leaks in water supplies [2, 3]. Moreover, health monitoring and connected medical devices will undoubtedly transform healthcare services.

Irrefutable benefits proposed by the IoT paradigm, however, are coupled with serious security issues. In fact, some manufacturers continue to sacrifice security concerns for the sake of reducing costs and increasing the speed of delivering IoT benefits, while consumers struggle to update firmware and change default user credentials. At the same time, attackers are taking advantage of vulnerable devices, maliciously manipulating Internet-wide deployed IoT nodes, causing a profound impact on the security and the resiliency of the entire Internet. We recently have witnessed various cyber attacks launched by IoT-specific malware, which demonstrated the severity of such exploited and coordinated IoT devices. In case of Mirai [4], the primary DNS provider in the US, Dyn, became the target of an orchestrated Denial of Service (DoS) attack, jeopardizing the profit and reputation of its clients. Further, poorly designed devices can expose user data to theft by leaving data streams inadequately protected [5].

Such and other incidents stimulated the research and cyber security operational communities to approach the IoT security issue. Nevertheless, many challenges remain unsolved, including the visibility and accessibility of IoT-specific data. Indeed, IoT devices are deployed in many private domains, access to which is restricted or even impossible. As a result, we observe a deficiency of techniques aiming to identify large-scale compromised IoT devices in a near real-time fashion. A thorough investigation of the generated traffic by compromised devices, however, would enable not only the inference of such devices but also the generation of malicious traits for preventing further exploitation. Given that the investigation of network telescope data proved its efficiency in observing Internet-wide unsolicited activities via executing passive measurements and analysis [6], we envisioned that such traffic would also generate valuable insights regarding large-scale illicit activities of IoT devices. Indeed, a network telescope, also known as a darknet, is a set of sensors which represent routable, allocated, yet unused IP addresses. The absence of Internet services associated with these IP addresses, render them an effective approach to amalgamate Internet-wide unsolicited events.

In this paper, we explore passive network measurements to comprehend the magnitude of IoT exploitations, differentiate their malicious activities, and analyze their hosting environments. We also conduct discussions with various entities to obtain IP allocation information, which further allows us to attribute the IoT exploitations per business sector (i.e., education, financial, manufacturing, etc.). Specifically, we frame the contributions of this work as follows:

- Executing a large-scale empirical characterization of unsolicited activities generated by IoT devices. To this end, we leverage network telescope (passive) measurements in combination with active measurements to infer compromised IoT devices and reveal the nature of their illicit activities. The generated intelligence aims at providing valuable insights that would aid in proper remediation.
- Generating amalgamated statistics regarding compromised devices and their hosting environments, including sector information, which has never been reported before.

The paper is organized as follows. Section II reviews the related work and emphasizes the contributions of this paper. Section III elaborates on the techniques which enable the inference of large-scale compromised IoT devices and their characterization. In Section IV, we shed light on the findings, including results related to sector-based IoT exploitations. Finally, Section V summarizes this paper and pinpoints few topics for future work.

II. Related Work

In this section, we review the literature by elaborating on three IoT-related topics, namely, data capturing, empirical measurements for device characterization and device fingerprinting.

Various recent research works have been devoted to developing IoT-centric honeypots aiming at gathering and analyzing IoT-specific security-related data. To this end, Pa et al. [7] pioneered a honeypot designed primarily for IoT to investigate ongoing attacks against Telnet services and discovered several malware families. Auxiliary, Guarnizo et al. [8] proposed the Scalable high-Interaction Honeypot (SIPHON) platform for the IoT paradigm and demonstrated how the combination of a limited number of physical devices and worldwide wormholes permit the emulation of numerous IoT devices on the Internet. This aimed at attracting massive IoT malicious traffic. Several other works, including [9, 10], further proposed honeypots aiming at detecting malicious activities targeting industrial control systems.

In the context of empirical measurements for device characterization, Angrishi [11] analyzed IoT-centric malware while Costin et al. [12] performed a large-scale static analysis of embedded firmware to explore IoT insecurities. Further, Fachkha et al. [13] leveraged passive measurements and analyzed attackers' intentions when targeting protocols of Internet-facing CPS. By monitoring requests to a network telescope and employing filters to distinguish Mirai traffic, Antonakakis et al. [14], in contrast, identified 1.2 million Mirai infected IP addresses associated with various deployment environments and types of IoT devices.

In an attempt to infer IoT-specific traits, a number of researchers pursued IoT device fingerprinting. For instance, Meidan et al. [15] classified IoT nodes connected to an organization's network by solely observing network traffic. Similarly, Formby et al. [16]

designed two approaches for device fingerprinting by leveraging observations rooted in cross-layer response times and unique physical properties of IoT devices.

This work compliments the available contributions by extending network telescope research to investigate malicious activities generated by Internet-scale IoT devices. To this end, we develop a data-driven approach to pinpoint compromised IoT devices, investigate and differentiate their unsolicited actions, and examine their hosting environments. Moreover, by conducting discussions with various realms which operate IoT devices, we attribute such identified IoT devices by their hosting sectors (i.e., education, financial, manufacturing, etc.); an initiative that has never been attempted (or reported) before.

III. Proposed Approach

In this work, we take a first step towards addressing the problem of IoT security at large. We approach this goal by correlating passive monitoring of Internet-wide network traffic with various datasets to discover compromised IoT devices in local realms and precisely determine their malicious traits. Figure 1 illustrates a holistic architecture of the proposed approach. We uniquely explore traffic collected by a network telescope (i.e., a set of routable, allocated, yet unused IP addresses). Characteristically, all traffic targeting this IP space is unsolicited [17]. Investigating network telescope traffic has recurrently demonstrated its effectiveness in inferring various types of malicious activities such as DDoS victims [18–20] and probing activities [21–24].

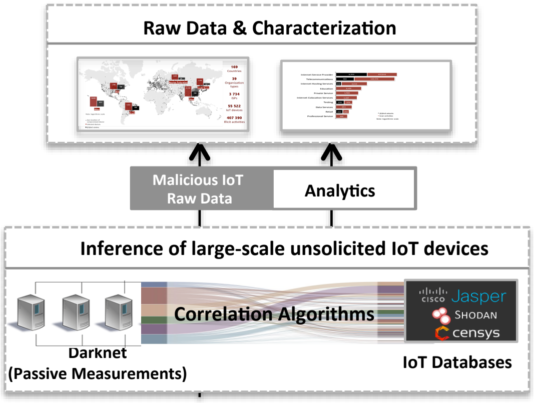
Raw Data & Characterization
Malicious IoT Raw Data
Analytics
Inference of large-scale unsolicited IoT devices
Correlation Algorithms
Darknet (Passive Measurements)
Jasper
SHODAN
censys
IoT Databases

Fig. 1. A data-driven approach for inferring and characterizing compromised IoT devices

The lacking of visibility related to IoT-specific data coupled with the shortage of IoT fingerprinting approaches indeed challenge the task of distinguishing IoT devices from other Internet hosts. To this end, we leverage the search engine Shodan [25] as a database of IoT devices. Specifically, we adopt its API to correlate malicious hosts (targeting the network telescope) and identify IoT devices.

We further enrich the generated intelligence with the hosting environment of such IoT exploitations, including country and ISP information. To this end, we correlate each IP address associated with the unsolicited IoT devices with internal and external databases. We utilized internal knowledge (gathered by conducting discussions with various Internet entities) rendered by IP ranges associated with various business sectors. Complementary, we employed MaxMind [26] for the remaining geolocation requirements.

IV. Empirical Evaluation of IoT Maliciousness

In this section, we provide a characterization of IoT maliciousness in terms of illicit activities and hosting environments. The executed analysis draws upon close to 1.2 TB of darknet data that was collected from a /8 network telescope provided by CAIDA [27] for a recent 24-hour period. We distinguish two classes of maliciousness presented in this period. These are *(i)* victims of DDoS attacks, including victims of TCP, UDP, and ICMP flooding; and *(ii)* hosts that conduct horizontal, vertical, and strobe scans against Internet hosts. Precisely, we identified close to 5,000 Internet hosts that have fallen victims of more than 30,000 DDoS attacks. We also identified nearly 1.2 million infected hosts, which generated 4.5 million scanning activities.

The correlation algorithm between network telescope traffic and the IoT dataset yielded nearly 56,000 IoT devices, which generated illicit Internet traffic, representing 5% of total inferred malicious activities. Auxiliary, nearly 940 IoT devices fell victims of 9,000

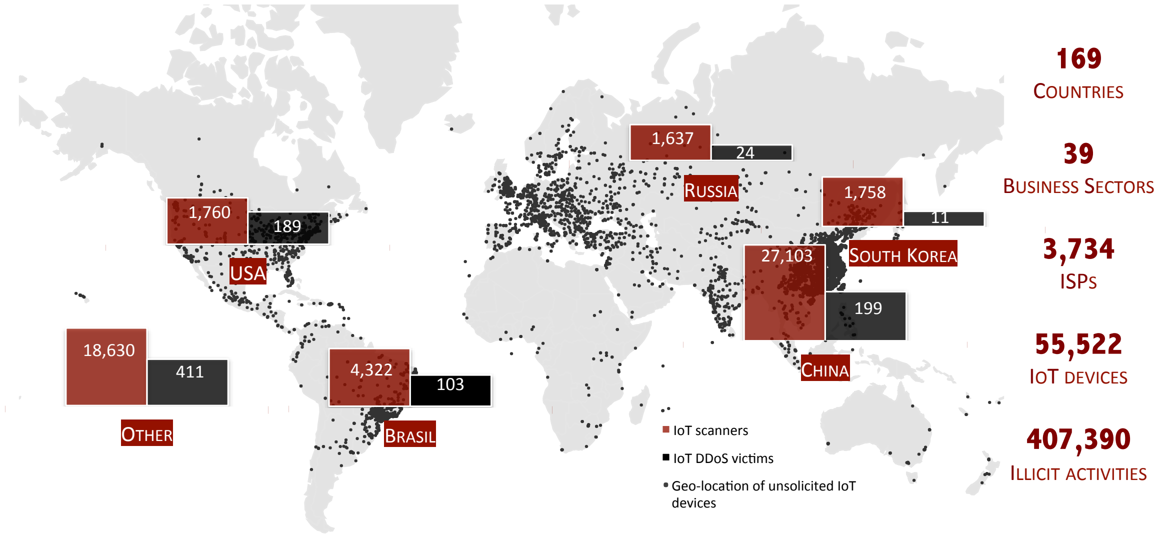

Fig. 2. IoT Devices: Global exploitations and DoS victims

DDoS attacks. It is worthy to pinpoint that IoT devices presented 19% of the total identified DDoS victims. In the same manner, 5% of the infected Internet hosts that attempted to explore other Internet hosts are IoT devices, which generated 9% of total scans. The latter is an alarming number of IoT malicious activities taking into account that we only analyzed one day of network telescope traffic.

We identified the presence of compromised IoT devices in 169 countries worldwide, hosted by 39 various business sectors, in nearly 4,000 ISPs. Figure 2 illustrates the global distribution of such unsolicited IoT devices and emphasizes the top 5 source countries. Specifically, we detected devices in China (49%), followed by Brazil (8%), United States (3%), South Korea (3%), and Russia (3%). In total, these countries hosted 66% of the affected devices, which generated close to 51% of the inferred illicit activities.

The significant number of IoT-generated malicious activities was found to be associated with various hosting sectors, such as Internet service providers (40%) and telecommunication entities (30%), which hosted 42% and 36% of compromised devices, respectively. Figure 3 illustrates the most affected hosting sectors and their corresponding number of misdemeanors. While the aim of the aforementioned IoT-specific malicious activities is unclear, the presence of such devices in educational, governmental and professional services could be benign (for research purposes). Additionally, despite the relatively low number of compromised devices in critical sectors such as manufacturing and financial, their presence in

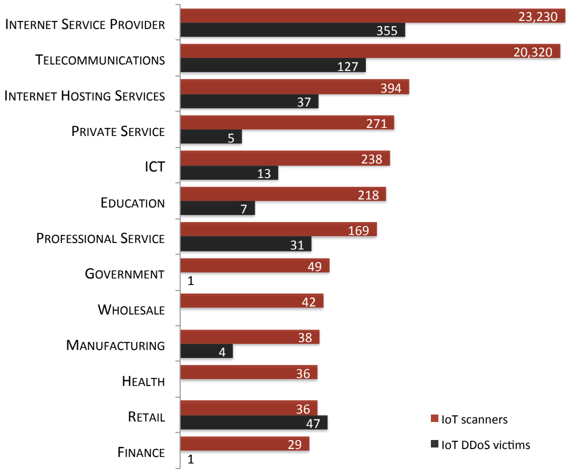

Fig. 3. Exploitations/victims by business sectors

TABLE I
TOP ISPS HOSTING THE MOST IOT EXPLOITATIONS

ISP	Devices	Scans	Scans/Device	ISP	Devices	Scans	Scans/Device
Vivo	3105	23662	7.62	China Telecom	1,556	7,332	4.71
China Mobile Guangdong	2,978	7,805	2.62	China Telecom Zhejiang	1,343	4,496	3.35
China Unicom Liaoning	2,818	10,757	3.82	China Telecom Sichuan	1,238	3,169	2.56
China Telecom Guangdong	2,017	6,091	3.02	China Telecom fujian	1,237	6,922	5.6
China Telecom jiangsu	1,569	5,628	3.59	China Telecom Hunan	905	3,870	4.28
Other ISPs	17,814	12,1661	6.83				

such sectors is significantly alarming and could cause serious issues, including exfiltration of sensitive data and environmental damages. Auxiliary, we observed unsolicited IoT devices hosted in the healthcare industry. The stolen information from such devices could cause momentous privacy breaches, fraudulent insurance claims, and more severely, such exploitations could threaten patients' lives.

A. IoT devices conducting network scans

In this section, we investigate the hosting environments of IoT devices which were found to be aggressively scanning the Internet space. We center our investigation around the 5 countries that hosted the highest number of IoT devices. The latter devices generated 50% of total scanning activities from around 66% of the total volume of affected IoT nodes. Table II summarizes the activities of such devices by country.

TABLE II
NUMBER OF INFECTED IOT DEVICES AND SCANNING ACTIVITIES BY COUNTRY

Country	Devices	Scans	Scans/Device
🇨🇳 China	27,103	98,444	3.63
🇧🇷 Brazil	4,322	38,516	8.91
🇺🇸 United States	1,760	21,106	11.99
🇰🇷 South Korea	1,758	29,436	16.74
🇷🇺 Russia	1,637	13,891	8.49
Other countries	18,630	197,347	10.59

The average rate of scanning activities per one compromised device indicates that devices which are located in South Korea and the US generate malicious traffic more aggressively than those which were found in China, the country with the highest number of compromised devices. Deliberate examination of this rate uncovered that the most aggressive scan activities are generated by few devices hosted by numerous business sectors, including, ISPs, the US government, health, education, and the financial sector (in particular banks). In South Korea, the majority of such illicit events are hosted by telecommunication companies and IPSs. The study of ISPs in the countries with the highest presence of unsolicited IoT devices uncovered that Vivo, the larger telecommunications company in Brazil, appears to be number one host of unsolicited IoT devices, presented by 6% of total compromised devices. Further, 5% of compromised devices are hosted by China Mobile Guangdong, China Unicom Liaoning and China Telecom Guangdong. Table I lists the top 10 ISPs which host the most IoT compromised devices.

B. IoT devices as DDoS victims

In this section, we investigate the hosting environments of IoT devices which have fallen victims of DDoS attacks. Such devices were identified as representing 63% of total number of inferred DDoS attacks in the aforementioned top countries. In fact, these attacks affected 56% of inferred devices. Table III specifies the number of IoT DDoS victims and attacks by country.

Please note that attack/device represents the average number of attacks per IoT device, and not the magnitude of such attacks. In this context, we observed that the devices in China attracted the highest number of illicit activities. Precisely, their number is twice higher than in other countries. A closer investigation of this fact unveiled that the significant

TABLE III
NUMBER OF IOT DDOS VICTIMS AND ATTACKS BY COUNTRY

Country	**Devices**	**Attacks**	**Attack/Device**
China	199	3,033	15.24
United States	189	1,579	8.35
Brazil	103	674	6.54
Russia	24	131	5.46
South Korea	11	54	4.91
Other countries	411	3,179	7.73

number of such devices is associated with 4 ISPs that have suffered persistent attacks. In fact, these ISPs, which are listed in Table IV, hosted 60% of IoT DDoS victims in China, absorbing close to 80% of attacks, as observed by the monitored network telescope.

TABLE IV
TOP 4 ISPS HOSTING IOT DDOS VICTIMS IN CHINA

ISP	**Devices**	**Attacks**	**Attack/Device**
China Telecom backbone network	75	1631	21.75
China Telecom Shanghai	18	332	18.44
China Telecom	16	308	19.25
China Telecom Shandong	10	162	16.2
Other ISPs	80	600	7.5

The study of ISPs in the aforementioned countries uncovered that the ISP which hosted a significant number of DDoS victims is SNH Servicos de Internet Ltda., which is an Internet provider in Brazil. The closest follower is China Telecom backbone network. Table V summarizes the top 5 ISPs that were found to be hosting the highest number of IoT DDoS victims.

TABLE V
ISPS HOSTING IOT DDOS VICTIMS IN COUNTRIES WITH THE HIGHEST NUMBER OF IOT EXPLOITATIONS

ISP	**Devices**	**Attacks**
SNH Servicos de Internet Ltda.	90	463
China Telecom backbone network	75	1,631
AT&T U-verse	40	641
Google	34	134
Amazon Technologies	25	119
Other ISPs	238	2,352

V. CONCLUDING REMARKS

In this effort to address the IoT security problem at large, we extended passive measurements and analysis by scrutinizing network telescope data to report on malicious activities generated by compromised IoT devices. We achieved our goal by thoroughly investigating a significant amount of network telescope data and by executing correlations auxiliary databases. In particular, through imperative discussions with numerous operators, we were able to obtain information about IP ranges belonging to various business sectors, permitting the capability to identify IoT exploitation in such sectors. Some of the outcome disclosed more than 407,000 illicit events, originating from nearly 56,000 unsolicited and malicious IoT devices. As for future work, we will be investigating the root cause of such IoT exploitations, including IoT-specific malware.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the anonymous reviewers for their constructive feedback. This work was partially supported by a grant from the U.S. National Science Foundation (NSF) (Office of Advanced Cyberinfrastructure (OAC) #1755179). The authors would also like to thank Dubai Electronic Security Agency (DESC) in Dubai, UAE for partially supporting this work.

REFERENCES

[1] I. Bisio, A. Delfino, F. Lavagetto, and A. Sciarrone, “Enabling IoT for In-Home Rehabilitation: Accelerometer Signals Classification Methods for Activity and Movement Recognition,” *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 135–146, Feb 2017.

[2] Inter-American Development Bank (IDB), in association with the Korea Research Institute for Human Settlements (KRIHS), “Smart cities - international case studies,” http://www.iadb.org/en/topics/emerging-and-sustainable-cities/international-case-studies-of-smart-cities,20271.html, accessed 2018-03-05.

[3] E. Bou-Harb, W. Lucia, N. Forti, S. Weerakkody, N. Ghani, and B. Sinopoli, “Cyber Meets Control: A Novel Federated Approach for Resilient CPS Leveraging Real Cyber Threat Intelligence,” *IEEE Communications Magazine*, vol. 55, no. 5, pp. 198–204, 2017.

[4] B. Herzberg, D. Bekerman, and I. Zeifman, “Breaking Down Mirai: An IoT DDoS Botnet Analysis,” https://www.incapsula.com/blog/malware-analysis-mirai-ddos-botnet.html, October 2016, accessed 2018-03-05.

[5] M. Stanislav and T. Beardsley, “HACKING IoT: A Case Study on Baby Monitor Exposures and Vulnerabilities,” *Rapid 7*, 2015.

[6] C. Fachkha and M. Debbabi, “Darknet as a source of cyber intelligence: Survey, taxonomy, and characterization,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1197–1227, 2016.

[7] Y. M. P. Pa, S. Suzuki, K. Yoshioka, T. Matsumoto, T. Kasama, and C. Rossow, “IoTPOT: A novel honeypot for revealing current IoT threats,” *Journal of Information Processing*, vol. 24, no. 3, pp. 522–533, 2016.

[8] J. D. Guarnizo, A. Tambe, S. S. Bhunia, M. Ochoa, N. O. Tippenhauer, A. Shabtai, and Y. Elovici, “SIPHON: Towards scalable high-interaction physical honeypots,” in *Proceedings of the 3rd ACM Workshop on Cyber-Physical System Security*. ACM, 2017, pp. 57–68.

[9] D. I. Buza, F. Juhász, G. Miru, M. Félegyházi, and T. Holczer, “CryPLH: Protecting smart energy systems from targeted attacks with a PLC honeypot,” in *International Workshop on Smart Grid Security*. Springer, 2014, pp. 181–192.

[10] E. Vasilomanolakis, S. Srinivasa, C. G. Cordero, and M. Mühlhäuser, “Multi-stage attack detection and signature generation with ICS honeypots,” in *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*, April 2016, pp. 1227–1232.

[11] K. Angrishi, “Turning Internet of Things (IoT) into Internet of Vulnerabilities (IoV): IoT Botnets,” *CoRR*, vol. abs/1702.03681, 2017, http://arxiv.org/abs/1702.03681.

[12] A. Costin, J. Zaddach, A. Francillon, D. Balzarotti, and S. Antipolis, “A large-scale analysis of the security of embedded firmwares,” in *USENIX Security Symposium*, 2014, pp. 95–110.

[13] C. Fachkha, E. Bou-Harb, A. Keliris, N. Memon, and M. Ahamad, “Internet-scale probing of CPS: Inference, characterization and orchestration analysis,” in *The Network and Distributed System Security Symposium (NDSS), To appear*, 2017.

[14] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis *et al.*, “Understanding the Mirai Botnet,” in *USENIX Security Symposium*, 2017.

[15] Y. Meidan, M. Bohadana, A. Shabtai, J. D. Guarnizo, M. Ochoa, N. O. Tippenhauer, and Y. Elovici, “ProfilIoT: A Machine Learning Approach for IoT Device Identification Based on Network Traffic Analysis,” in *Proceedings of the Symposium on Applied Computing*, ser. SAC ’17. New York, NY, USA: ACM, 2017, pp. 506–509.

[16] D. Formby, P. Srinivasan, A. Leonard, J. Rogers, and R. Beyah, “Who’s in control of your control system? Device fingerprinting for cyber-physical systems,” in *Network and Distributed System Security Symposium (NDSS)*, 2016.

[17] E. Bou-Harb, M. Debbabi, and C. Assi, “A novel cyber security capability: Inferring internet-scale infections by correlating malware and probing activities,” *Computer Networks*, vol. 94, pp. 327–343, 2016.

[18] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage, “Inferring internet denial-of-service activity,” *ACM Transactions on Computer Systems (TOCS)*, vol. 24, no. 2, pp. 115–139, 2006.

[19] C. Fachkha, E. Bou-Harb, and M. Debbabi, “Inferring distributed reflection denial of service attacks from darknet,” *Computer Communications*, vol. 62, pp. 59–71, 2015.

[20] ——, “On the inference and prediction of DDoS campaigns,” *Wireless Communications and Mobile Computing*, vol. 15, no. 6, pp. 1066–1078, 2015.

[21] A. Dainotti, A. King, K. Claffy, F. Papale, and A. Pescapé, “Analysis of a /0 stealth scan from a botnet,” *IEEE/ACM Transactions on Networking (TON)*, vol. 23, no. 2, pp. 341–354, 2015.

[22] E. Bou-Harb, M. Debbabi, and C. Assi, “A statistical approach for fingerprinting probing activities,” in *Availability, Reliability and Security (ARES), 2013 Eighth International Conference on*. IEEE, 2013, pp. 21–30.

[23] ——, “Behavioral analytics for inferring large-scale orchestrated probing events,” in *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2014, pp. 506–511.

[24] E. Bou-Harb, N.-E. Lakhdari, H. Binsalleeh, and M. Debbabi, “Multidimensional investigation of source port 0 probing,” *Digital Investigation*, vol. 11, pp. S114–S123, 2014.

[25] J. Materly, “Shodan,” https://shodan.io, 2009, accessed 2018-03-05.

[26] MaxMind, Inc., “GeoIP2 Databases,” https://www.maxmind.com/en/geoip2-databases, accessed 2018-03-05.

[27] “UCSD Network Telescope – Near-Real-Time Network Telescope Dataset,” http://www.caida.org/data/passive/telescope-near-real-time_dataset.xml, accessed 2018-03-05.