ACCEPTED FROM OPEN CALL

Internet of Malicious Things: Correlating Active and Passive Measurements for Inferring and Characterizing Internet-Scale Unsolicited IoT Devices

Farooq Shaikh, Elias Bou-Harb, Nataliia Neshenko, Andrea P. Wright, and Nasir Ghani

The authors describe a first step toward exploring the utilization of passive measurements in combination with the results of active measurements to shed light on the Internet-scale insecurities of the IoT paradigm. By correlating results of Internet-wide scanning with Internet background radiation traffic, they disclose close to 14,000 compromised IoT devices in diverse sectors, including critical infrastructure and smart home appliances.

Abstract

Advancements in computing, communication, and sensing technologies are making it possible to embed, control, and gather vital information from tiny devices that are being deployed and utilized in practically every aspect of our modernized society. From smart home appliances to municipal water and electric industrial facilities to our everyday work environments, the next Internet frontier, dubbed IoT, is promising to revolutionize our lives and tackle some of our nations' most pressing challenges. While the seamless interconnection of IoT devices with the physical realm is envisioned to bring a plethora of critical improvements in many aspects and diverse domains, it will undoubtedly pave the way for attackers that will target and exploit such devices, threatening the integrity of their data and the reliability of critical infrastructure. Further, such compromised devices will undeniably be leveraged as the next generation of botnets, given their increased processing capabilities and abundant bandwidth. While several demonstrations exist in the literature describing the exploitation procedures of a number of IoT devices, the up-to-date inference, characterization, and analysis of unsolicited IoT devices that are currently deployed "in the wild" is still in its infancy. In this article, we address this imperative task by leveraging active and passive measurements to report on unsolicited Internet-scale IoT devices. This work describes a first step toward exploring the utilization of passive measurements in combination with the results of active measurements to shed light on the Internet-scale insecurities of the IoT paradigm. By correlating results of Internet-wide scanning with Internet background radiation traffic, we disclose close to 14,000 compromised IoT devices in diverse sectors, including critical infrastructure and smart home appliances. To this end, we also analyze their generated traffic to create effective mitigation signatures that could be deployed in local IoT realms. To support large-scale empirical data analytics in the context of IoT, we make available the inferred and extracted IoT malicious raw data through an authenticated front-end service. The outcomes of this work confirm the existence of such compromised devices on an Internet scale, while the generated inferences and insights are postulated to be employed for inferring other similarly compromised IoT devices, in addition to contributing to IoT cyber security situational awareness.

Introduction

The Internet of Things (IoT) paradigm refers to scenarios where network connectivity and computing capability extend to embedded sensors, allowing these devices to generate, exchange, and consume data with minimal human intervention [1]. This paradigm is being realized and facilitated by critical advancements in computing power, electronics miniaturization, and network interconnections. Indeed, the large-scale deployment of IoT devices promises to transform many aspects of our contemporary lives. Innovative IoT products such as Internet-enabled appliances, home automation components, and energy management devices are creating the vision of the *smart home*, offering more security and energy efficiency. Other personal IoT devices such as wearable fitness, health monitoring devices, and network-enabled medical devices are transforming the way healthcare services are being delivered. IoT systems in other domains such as networked vehicles, intelligent traffic systems, and embedded sensors at roads and bridges are redefining the idea of *smart cities*, which help minimize congestion and energy consumption. Additionally, IoT technology offers the possibility to transform agriculture, and energy production and distribution by increasing the availability of information along the value chain of production using networked distributed sensors.

While the deployment of such IoT devices and paradigms in various domains will introduce a plethora of benefits and advantages, their unique attributes coupled with their interconnected nature will indeed present new security challenges. For instance, the competitive cost and technical constraints on IoT devices will challenge manufacturers to adequately design security features into these devices, potentially creating security and long-term maintainability vul-

Digital Object Identifier:
10.1109/MCOM.2018.1700685

Please submit all author affiliations.

nerabilities. Along with potential security design deficiencies, the sheer increase in the number, type, and nature of IoT devices would undeniably increase the opportunities for exploitation. Vulnerable IoT devices could serve as entry points for attackers, allowing the malicious re-programming of a device or causing it to malfunction. Further, poorly designed devices can expose user data to theft by leaving data streams inadequately protected. Failing or malfunctioning devices can also create a number of serious security vulnerabilities. When coupled with the highly interconnected nature of IoT devices, every poorly secured Internet-connected device would potentially affect the security and resilience of the Internet on a global scale, not just locally. For example, an unprotected refrigerator or television in the United States that is infected by malware might send thousands of harmful spam emails to worldwide recipients using the owner's home WiFi Internet connection. Additionally, exploited Internet-scale IoT devices could be leveraged by an attacker to orchestrate malicious botnets, causing momentous damage to corporate and Internet services by means of denial of service (DoS) attacks and other typical misdemeanors. Examples of the latter already exist, with the rise of the Mirai malware that exploited Internet-scale IoT cameras to cause drastic damage to Internet infrastructure services. Additionally, very recently, the Hajime IoT worm [2], which extends Mirai with more harmful capabilities, is indeed defining a new era of malicious bots capable of hindering the integrity of critical infrastructure and leaking private information.

To this end, the imperative tasks of quantifying, characterizing, and attributing such vulnerable IoT devices would render a first step toward uncovering their inherited vulnerabilities and understanding their malicious intents. Thus, in this article, we exploit the results of active measurements through Internet-wide scanning in conjunction with passive measurements in the context of darknet traffic analysis to shed light on such devices and analyze their unsolicited network traffic characteristics. Specifically, we frame the article's contributions in the following three threads:

•Proposing and evaluating an innovative approach to infer, characterize, and attribute unsolicited Internet-scale IoT devices by correlating the results of passive and active empirical measurements. To the best of our knowledge, this work is among the first to explore such an approach, which addresses the IoT paradigm.

•Generating IoT-specific malicious signatures by scrutinizing passive measurements. Such signatures, which are based on fuzzy hashes, are envisioned to be employed for deployments in local IoT realms for effective mitigation as well as to infer other Internet-wide unsolicited IoT devices. As noted in [3], IoT-specific empirical attack signatures currently do not exist, rendering the proposed approach highly impactful and operationally very beneficial.

•Reporting on close to 14,000 compromised IoT devices related to smart home appliances, critical infrastructure, and automated control sectors. In this context, we generate amalgamated statistics related to these inferred and exploited IoT devices, including their hosting environments. All the generated inferences, including IoT malicious raw data, are made publicly accessible through an authenticated service. We postulate that such generated cyber threat intelligence could be exploited, in near real time, for effective IoT cyber security situational awareness, notification, and remediation. The road map of this article is as follows. In the next section, we review a number of related works in the context of IoT security. Following that, we elaborate on the proposed approach, and detail its rationale and employed techniques. Then we present the empirical evaluations and analyze the obtained results. The final section provides concluding remarks and pinpoints several insightful topics that pave the way for future work in this impactful IoT security research area.

Related Work

In this section, we review a number of research efforts in various related topics, namely, IoT context-aware permission models, IoT security and protocol analysis, and passive and active measurements for device characterization and vulnerability analysis.

The majority of IoT security research work has been dedicated to synthesizing IoT context-aware permission models. Such research endeavors address the problem of creating and enforcing collaborative models to secure IoT environments from malicious scenarios. For instance, Yu *et al.* [3] proposed a policy abstraction language that is capable of capturing relevant environmental IoT contexts, security-relevant details, and cross-device interactions to vet IoT-specific network activities. Further, the authors proposed a crowd-sourced repository where IoT operators can share derived attack signatures, which deviate from the captured benign policies. To provide enforcement capabilities in various IoT realms, the authors extended notions related to software-defined networking (SDNs) to the IoT context by proposing the utilization of micro-middleboxes that could provide real-time remediation to possibly vulnerable devices. Along the same research direction, Jia *et al.* [4] proposed ContextIoT, a system that is capable of supporting complex IoT-relevant permission models through performing program-flow and runtime taint analysis. The authors prototyped the system using the Samsung SmartThings platform and demonstrated that ContexIoT app patching adds nearly negligible delay, and the permission request frequency is far below the threshold that is considered to trigger user annoyance. In another closely related work, Fernandes *et al.* [5] proposed an approach that aims to restrict generated traffic flows from an exploited IoT application. The approach is based on taint arithmetic, which initially tracks an application's program flow to subsequently flag policy violations. Indeed, the aforementioned research contributions address IoT security within certain IoT realms. In contrast, in this work, we aim at providing an Internet-scale perspective on IoT maliciousness. To this end, we uniquely analyze and correlate empirical data to:

- Infer such unsolicited IoT devices in addition to notifying concerned IoT realms of such exploitations
- Generate IoT-specific malicious signatures that could be used for effective mitigation

Along with potential security design deficiencies, the sheer increase in the number, type, and nature of IoT devices would undeniably increase the opportunities for exploitation. Vulnerable IoT devices could serve as entry points for attackers by allowing the malicious re-programming of a device or causing it to malfunction.

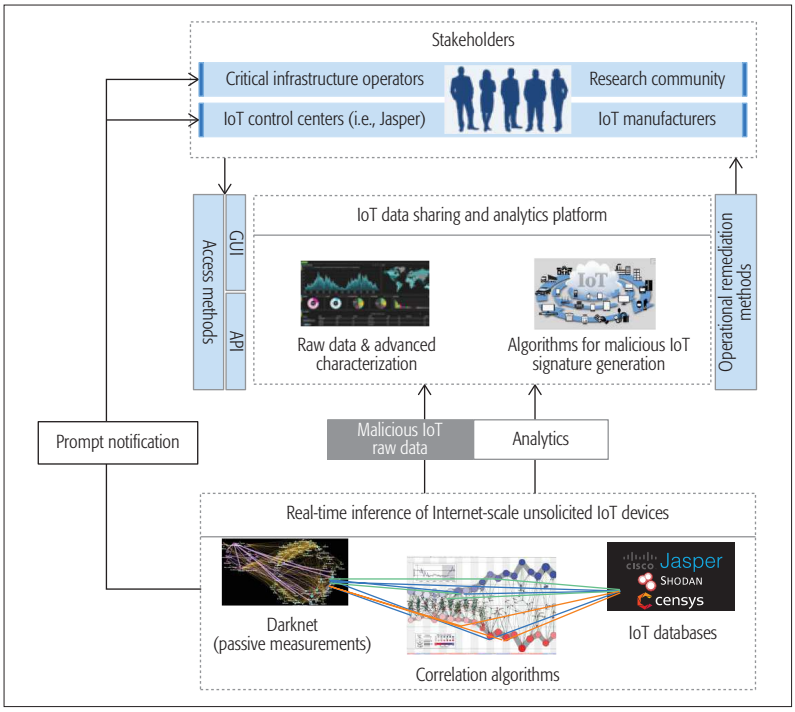

Figure 1. Inferring and mitigating Internet-scale unsolicited IoT devices: a network telescope approach.

Several other research endeavors have analyzed specific IoT security issues and protocol weaknesses. For instance, Ur *et al.* [6] addressed IoT access control by studying numerous types of home automation devices. By thoroughly investigating their ownership rules, roles, and monitoring capabilities, the authors unveiled various insights to shed light on usable access control. Their conducted study noted that some IoT devices allowed physical access control, which overrides the already applied access models, threatening the pairing security of such devices. A major research gap that was extracted from this work was related to the lack of mechanisms for provisioning auditing/monitoring of such IoT devices, rendering it impossible for users to identify who has accessed their devices and what types of activities were generated. In an alternative work, Ronen and Shamir [7] demonstrated information leakage attacks by instrumenting a set of IoT smart lights. To this end, the authors initially exploited protocol weaknesses to gain access to the connected local network and subsequently exfiltrated sensitive data from an air-gapped office building. The authors also extended their attacks by forcing the lights to act in an orchestrated fashion, generating a frequency that aims to cause seizures in people who have photosensitive epilepsy. In a similar fashion, Ho *et al.* [8] investigated state consistency and unwanted unlocking attacks by exploring protocol and system vulnerabilities in IoT smart locks. The authors demonstrated how trust models, network designs, and replay activities could be instrumented to cause security issues related to the revocation procedures of such locks in addition to forcing secure locks to be accidentally unlocked. The latter indeed paves the way for the occurrence of physical threats. Remediation methods based on secure architectures and cryptographic primitives were also proposed by the authors to mitigate the presented attack scenarios. In contrary, in this work, we propose and evaluate an operational cyber security capability that is rendered by inferring Internet-scale, heterogeneous, and unsolicited IoT devices that are deployed in various realms by leveraging empirical measurements and data. To this end, we report on such IoT devices to generate prompt cyber threat intelligence that could be leveraged for immediate remediation.

In the area of empirical measurements for device characterization and vulnerability analysis, Cui and Stolfo [9] executed a large-scale active probing of the Internet space to uncover close to 0.5 million vulnerable embedded devices. In more recent works, Costin *et al.* [10] statically analyzed more than 30,000 firmware images derived from embedded devices to shed the light on their insecurity, while Fachkha *et al.* [11] conducted passive measurements to analyze attackers' intentions when targeting protocols of Internet-facing cyber-physical systems. The latter approach is quite similar to [12], where Bodenheim *et al.* evaluated the Shodan service (https://www.shodan.io/), a search engine for Internet-connected devices, in its capability to scan and index online industrial control systems. In contrast, in this work, we fuse the results of both passive and active empirical Internet traffic measurements to highlight on the maliciousness of Internet-scale IoT devices. Additionally, we uniquely generate IoT-specific attack signatures that could be used for effective remediation in local IoT realms. To support large-scale empirical data analytics, forensic investigations, and cyber security situational awareness in the context of IoT, we also make the generated inferences, insights, and extracted IoT unsolicited raw data publicly accessible to the research and operation communities at large through an online front-end service.

Proposed Approach

In this section, we detail the proposed approach and describe its aims, employed methods, and techniques. The proposed scheme is holistically illustrated in Fig. 1. In a nutshell, the approach endeavors to generate actionable cyber threat intelligence related to Internet-scale IoT devices by offering several data-driven methodologies, which mainly operate by scrutinizing passive empirical measurements. Such insights and inferences are postulated to be distributed, in an operational cyber security fashion, to IoT stakeholders (i.e., operators, manufacturers, etc.) for prompt remediation and thus mitigation. Furthermore, an artifact of the envisioned approach is a repository aimed at indexing malicious IoT empirical threat information (i.e., raw data and attack signatures) to be shared with the research and operational communities at large, hence facilitating advanced empirical IoT analytics as well as supporting further forensic investigations. In the sequel, we detail and elaborate on the components of the proposed approach.

Inferring and Characterizing Internet-Scale Compromised IoT Devices

Given the large-scale deployments of IoT devices in various worldwide environments, in addition to the variety of such devices in terms of their types and versions, it becomes quite challenging to generate effective insights that could be used

for inferring unsolicited infected IoT devices. This challenge is further complicated by the lack of data visibility into local IoT realms due to logistic and privacy concerns, in addition to the overall scarcity of malicious empirical data related to IoT devices [3]. Nevertheless, research and development attempts that endeavor to infer and characterize such IoT devices would be highly impactful and beneficial, as this will promptly pinpoint such devices to provide rapid remediation strategies. This, in turn, is aimed at thwarting the various malicious misdemeanors that could be generated by the constructed botnets, which are composed of such infected IoT devices.

In this context, we put forward a macroscopic approach to infer Internet-wide malicious IoT devices and realize it by correlating the results of empirical passive and active Internet measurements. We detail below such an approach and its rationale.

Exploiting Network Telescopes: Network telescopes, most commonly known as darknets [13], constitute a set of routable, allocated, but unused IP addresses. Such dark IP addresses render a tactical passive measurement approach to capturing Internet-scale maliciousness; since these IP addresses do not operate legitimate services, any traffic targeting them is indeed considered to be unsolicited. Figure 2 illustrates a common darknet architecture. Indeed, darknets are commonly distributed on specific Internet IP subspaces, which are often operated by Internet service providers (ISPs), educational and research entities, and corporate and backbone networks. Darknet IP addresses are, by nature, indistinguishable from other routable addresses, rendering them an effective technique to amalgamate Internet-wide, one-way unsolicited network traffic. In this work, we exploit network telescopes to identify network traffic originating from unsolicited IoT devices. The rationale here, as illustrated in Fig. 2, is rendered by our initial empirical observations, which concur that, compared to typical Internet hosts/machines, exploited IoT devices will also attempt to either propagate to infect other Internet IoT devices by launching scanning activities toward the Internet space, or fall victim to distributed DoS (DDoS) attacks. In either case, depending on the vantage points of the employed network telescopes, a varied portion of such activities will indeed target such dark IP spaces. To this end, well-established algorithms, methods, and techniques could be leveraged, which scrutinize darknet data (also known as Internet background radiation, IBR) to fingerprint such activities and thus infer sources of unsolicited IoT devices in addition to extracting their corresponding darknet traffic traces. For instance, such algorithms could be based on threshold analysis to infer IoT-generated scanning activities or exploit backscattered packet analysis (i.e., the analysis of reply packets originating from victims of DDoS attacks that were targeted by spoofed attackers) to pinpoint IoT devices that have been targeted by DDoS attacks [13]. In this work, to successfully recognize DDoS activities targeted IoT devices, we analyze close to 1.5 TB of real darknet data obtained from a /8 network telescope, which is facilitated through the Center for Applied Internet Data Analysis (CAIDA) [14]. It is worth mentioning that the proposed approach can only infer unsolicited IoT devices that generate traffic targeting the darknet IP space. This is indeed a typical limitation of any approach that exploits passive measurements by scrutinizing darknet data.

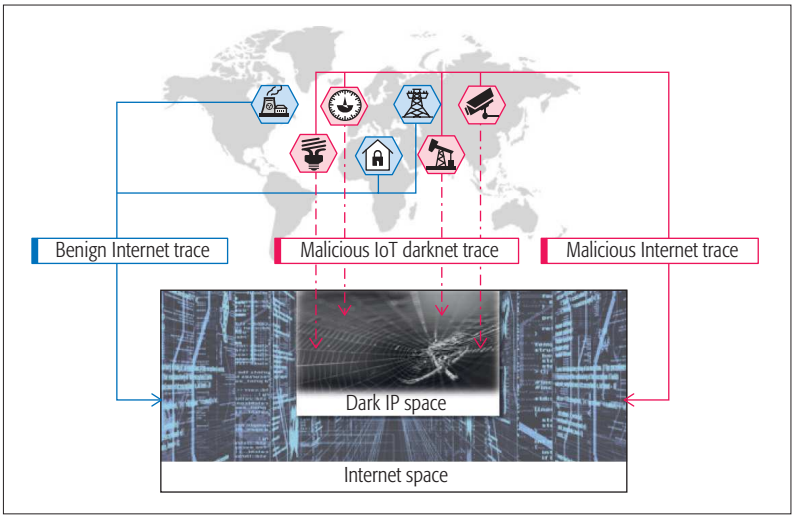

Figure 2. Network telescopes capturing Internet-scale IoT unsolicited traffic.

Leveraging Internet-Wide Scanning: While having a large empirical dataset characterizing Internet-scale unsolicited traffic is highly beneficial, there is now a need to filter those captured network sessions that are specifically generated by IoT devices. Indeed, the task of fingerprinting diverse types of IoT devices by solely analyzing IBR is currently an open research problem. To this end, in this work, we rely on correlating packet information, mainly source IP address and transport protocol, from previously extracted darknet sessions with the results of active measurements defined by Internet-wide scanning. For efficiency purposes, and given that the proposed approach is envisioned to provide near-real-time operational cyber security capability, we choose to leverage searchable databases of active measurements that are provided by Shodan and Censys (https://censys.io/) rather than performing our own Internet-wide scanning. The latter services continuously scan the Internet IP space looking for active Internet hosts, including IoT devices, while Shodan only indexes results related to IoT devices. We gather 275,478 online IoT IP addresses related to various types, including, home automation devices (i.e., thermostats, DVRs, etc.), IoT cameras, and industrial control IoT devices. The Shodan application programming interface (API) was leveraged to achieve this task. To select those darknet sessions that have been specifically generated by unsolicited/compromised IoT devices, we designed and implemented a correlation algorithm that fuses header information from passive measurements with the obtained IP addresses from active measurements. More specifically, given a list of source IP addresses extracted from darknet traffic and a list of IoT-specific IP addresses extracted from the IoT databases, the correlation algorithm performs linear search to infer a match. Please note that the current proposed approach operates in an offline mode. Future work will explore auxiliary online approaches to permit the near-real-time inference of Internet-scale compromised devices.

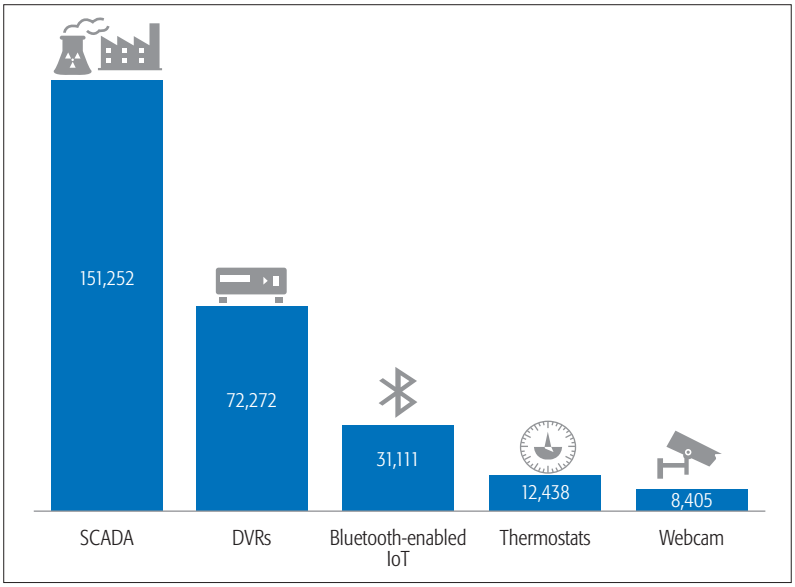

Figure 3. Distribution of investigated IoT devices by type.

Indeed, such correlation faces several challenges. First, in terms of darknet data, there is a need to sanitize such data to filter out misconfiguration traffic. Such traffic is the result of software, hardware, or routing errors that direct erroneous packets to the darknet. Second, in terms of the Shodan IoT database, one obstacle is the strenuous process of identifying relevant IoT information (i.e., types, IP addresses, etc.) and properly curating (download, sanitize, and store) the obtained information. Third, in terms of the actual correlation procedure, the design of the algorithm that executes the correlation should be optimized to perform efficient searches.

Generating IoT-Specific Malicious Signatures

Currently, there is a substantial lack of tangible malicious empirical indicators in the context of IoT [3]. This is mainly due to physical and logistic constraints that are strictly enforced by IoT operators in various realms. Additionally, data gathering and analytics efforts, aimed at addressing the distributed and heterogeneous nature of the IoT paradigm, are still in their infancy. Along these lines, it becomes highly imperative to generate such notions, which could aid in various IoT cyber security research and development endeavors. Indeed, the proposed approach was specifically designed in a manner to facilitate the generation of such artifacts. In particular, by executing network traffic correlations between the results of passive and active Internet measurements, we obtain access to rare unsolicited network traffic traces originating from Internet-scale compromised IoT devices. To this end, one imperative output would be to generate malicious IoT signatures, which characterize such extracted traces. These signatures are envisioned to be:

- Employed on newly incoming darknet sessions to fingerprint unsolicited IoT devices that have not been previously indexed by databases such as Shodan and Censys
- Distributed to local IoT realms where they can be deployed in traditional intrusion detection systems (IDSs) to support future proactive IoT inference and mitigation

To mutually support these two objectives, in this work, we exploit the concept of fuzzy hashing through tailoring and applying the Context Triggered Piecewise Hashing (CTPH) algorithm [15] on darknet traces that have been generated by the inferred unsolicited IoT devices. In particular, the IP header information is utilized in this context. The CTPH technique is advantageous in comparison with typical hashing, as it can provide a percentage of similarity rather than solely providing a binary output. This is particularly beneficial in the context of the first objective, where we apply the generated signatures on darknet traffic traces to verify if they possess some degree of similarity in comparison to previously obtained traces per the approach proposed above.

The CTPH algorithm operates only in the current context of the input, maintaining its state based solely on the last few bytes of the data file, thus producing a pseudo-random value as output. Essentially, the algorithm generates discrete hashes by dividing the file into multiple segments and computing hashes for these segments instead of computing a single hash for the entire file. In this way, localized segment changes do not affect the hashes for the rest of the file, and a degree of similarity can be determined in the case of almost identical files. The CTPH algorithm employs a rolling hash technique based on Alder32 checksum [15], which is computed for each data byte in the concerned file. This process is continuously iterated until all the bytes of the input file have been processed to generate the final signature. Readers who are interested in more details related to the CTPH algorithm are referred to [15].

Sharing of IoT Unsolicited Empirical Threat Information

Given the lack of real, empirical IoT threat information, we thought it would be of added value to make the extracted threat intelligence from the proposed approach publicly accessible to the research and operation communities at large. As shown in Fig. 1, we are currently developing an IoT sharing facility where IoT stakeholders (i.e., operators, researchers, etc.) can have access to (i) raw IoT unsolicited traffic traces to support large-scale IoT data analytics by leveraging such rare empirical data and (ii) the generated signatures to allow further forensic investigations as well as be employed at local realms for proactive inference and mitigation. To this end, we currently index the obtained insights, including near-real-time information related to Internet-scale compromised IoT devices coupled with their geolocation information as well as the generated signatures, in an accessible database, which is available through http://23.100.24.229/IoT-Sec/. This front-end service requires authentication; thus, we invite interested parties to contact the authors to gain access to such information.

Empirical Evaluation

In this section, we elaborate on the extracted inferences and insights that have been generated by employing the approach proposed above. We exploit passive Internet-scale darknet data in conjunction with active measurements provided by Shodan and Censys to infer the presence of malicious IoT devices, which have been deployed in consumer realms as well as in critical systems.

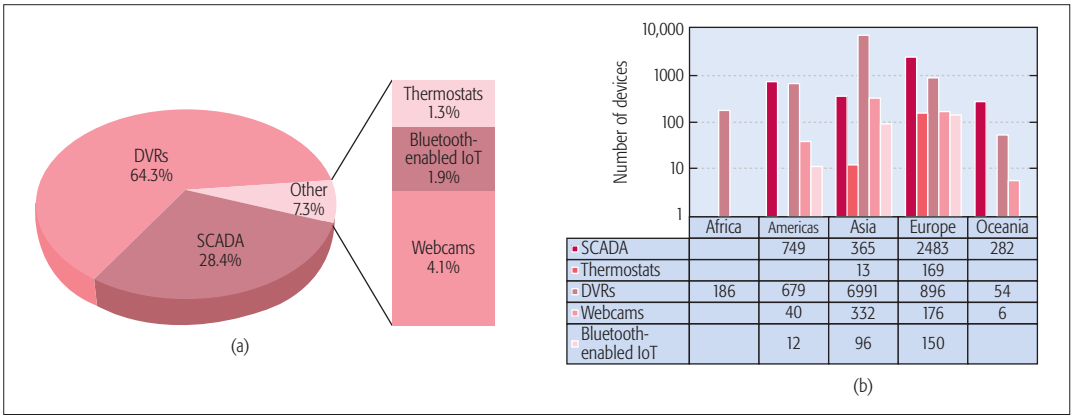

Figure 4. Distribution of exploited IoT devices: a) by types; b) by type and region.

The initial priorities of IoT vendors have been focused on providing novel functionality, getting products to market, and making IoT devices more accessible and easier to use. Unfortunately, security concerns have not received as much attention.

In total, we analyzed close to 1.5 TB of darknet data to verify the infection status of 275,478 IoT devices. The darknet data analyzed was for the months of November and December 2016 as well as January and February 2017, whereas the IoT IP addresses were retrieved from Shodan and Censys for the month of December 2016. For the sake of this work, we have investigated five categories of IoT devices, including digital video recorders (DVRs), webcams, thermostats, and Bluetooth-enabled devices, in addition to IoT sensors deployed deep in control automation systems. Figure 3 illustrates the distribution of analyzed devices within their corresponding categories. Overall, we extracted close to 165,000 IoT devices deployed in various supervisory control and data acquisition (SCADA) environments. This category represents 55 percent of the extracted devices. DVRs and Bluetooth-enabled IoT devices were less present, covering close to 26 and 11 percent of the total number of devices, respectively. The remaining 8 percent of investigated devices were IoT webcams and thermostats.

We proceeded by executing the correlation algorithm between passive and active measurements, as described above. The results unveiled the presence of almost 14,000 Internet-scale, currently active compromised IoT devices. We also note that close to 20 percent of such IoT devices were indeed found to be victims of DDoS attacks compared to those that have been inferred as generating scanning activities. Indeed, this outcome validates our intuition that there are a significant number of IoT devices which are involved in malicious activities in addition to corroborating our hypothesis that exploring network telescope traffic is an effective methodology to shed light on Internet-scale IoT maliciousness.

We proceeded by characterizing such exploited IoT devices as illustrated in Fig. 4a. We observed that DVRs occupied a significant portion (64.3 percent) of compromised IoT devices, a result which corroborates the exploitation of such devices earlier this year by the Mirai malware to launch debilitating DDoS attacks on Dyn DNS servers, which paralyzed part of the Internet infrastructure. IoT devices deployed in SCADA realms (mostly belonging to building automation systems, power utilities, and manufacturing plants) occupied 28.4 percent of the total share of unsolicited devices found in the darknet. IoT webcams, thermostats, and other Bluetooth-enabled devices were also shown to be compromised. Indeed, such inferences could promptly be leveraged by the operational cyber security community (i.e., IoT operators and manufacturers, cyber situational response teams, etc.) to aid in the rapid notification and thus mitigation of such exploitations.

To characterize the hosting environments of such IoT exploitations, we executed geo-location procedures by employing the MaxMind GeoIP2 database (https://www.maxmind.com/en/geoip2-databases). Our analysis revealed that the majority of compromised IoT devices are located in Asia followed by Europe and the Americas, as depicted in Fig. 4b. Asia was found to host a significant number of malicious DVRs, while exploited IoT devices deployed in control automation realms were mostly deployed in Europe and the Americas. Webcams and thermostats had a significantly smaller share when compared to DVRs and SCADA IoT devices, and were distributed between Asia and Europe. Figure 5 depicts the worldwide distribution of top countries hosting exploited IoT devices. Intuitively, this outcome is affected by the selected and investigated IoT IP addresses, the specific darknet data sample that has been utilized, and the timeframe of the executed analysis. We also generate further, more precise geo-location information such as hosting organization and ISPs, although we do not expose such results for sensitivity reasons.

We also generate IoT-specific malicious signatures per the proposed approach above by executing the CTPH algorithm on extracted IP header information from darknet traffic related to compromised IoT devices. To successfully achieve this, we leverage an open source implementation of the algorithm, namely, the ssdeep utility (http://ssdeep.sourceforge.net/). A sample of such signatures, related to four different IoT devices, for proof-of-concept purposes is summarized in Table 1. Such signatures render the first attempt ever to capture notions of IoT maliciousness by scrutinizing empirical data. Please recall that these signatures are postulated to be deployed on newly incoming darknet traces to fingerprint newly exploited IoT devices that have not been previously indexed by certain databases such as Shodan and Censys. Moreover, these signatures could be employed at local IoT realms to aid in the mitigation and thus remediation objectives.

From an operational cyber security perspective, we continue our work on developing and operating the proposed methodologies in real time to index and share the obtained insights with the research and security communities at large, to strongly support and facilitate IoT security research using empirical data, and to aid in the IoT remediation objective at scale.

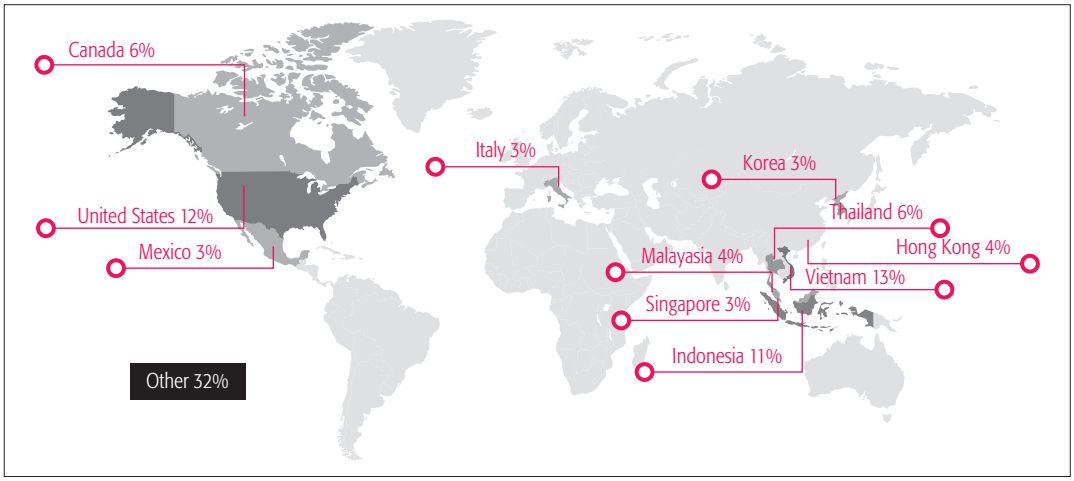

Figure 5. Global distribution of exploited IoT devices.

Device type	Signature
DVR	12288:wsIGM8PFc6fXPWW4cVsBK0GT5gkLXI5aurnz9k/Kk6:wsIP8i6fPWW4cmBKrT5gkzIxrnz9+Kk6
Webcam	3072:6OA062aJtmzOTYfpTYJ7JaZgVx3BAaTZQzTwcht79+8R+TMWs9Zm2g0ivLJ1p/jR:rgFmQyTEJaQmzTwMl982g0YF1IlYaJ
Printer	24576:XH9m8fEgLoZ7EqC0kf7tzH2uF/SD6dcZwEmGOqzH9m8fEgLf:4
Thermostat	6144:cMKa4Umz8VNPTg80mL4STGDs3+5FIwnVTF3gjGzTkpB/JkmBlDRkXY574OM231PL:0V80fJow3glAJkmQ23I8eme/X4AMG6Bb

Table 1. Signatures for fingerprinting unsolicited IoT devices.

Concluding Remarks and Future Directions

The Internet of Things is an emerging paradigm of technical, social, and economic significance. Projections for the impact of IoT on the Internet and economy are impressive, with a plethora of enterprises and analysts anticipating billions of connected IoT devices and a global economic impact of more than $11 trillion by 2025. While IoT deployments in the consumer sector have been receiving much hype, their corresponding implementations in critical system settings will undoubtedly provide massive benefits in terms of increased efficiency and cost reduction. Nevertheless, the initial priorities of IoT vendors have been focused on providing novel functionality, getting products to market, and making IoT devices more accessible and easier to use. Unfortunately, security concerns have not received as much attention. Indeed, the premise of this article is motivated by the lack of IoT-relevant empirical data, the shortage of IoT-specific malicious signatures, and the absence of operational IoT-centric cyber-infrastructure. To this end, this article leverages a significant amount of real, Internet-scale network telescope traffic coupled with the results of active measurements to shed light on IoT maliciousness. Correlation algorithms and methodologies are developed and executed to infer, characterize, and generate valuable malicious signatures related to unsolicited and diverse types of IoT devices that have been deployed in consumer and critical system sectors.

This work indeed presents a solid foundation on which future research efforts in this imperative IoT empirical security research area are currently being planned and pursued. Foremost, a large-scale, more thorough, empirical characterization ought to be conducted to deeply comprehend the severity and magnitude of Internet-wide IoT maliciousness. Additionally, further empirical analysis will be executed to understand the nature of the generated traffic of such compromised IoT devices. Data analytics rooted in machine and deep learning is also planned to be explored, to provide better characterization and usage of the generated IoT malicious signatures. From an operational cyber security perspective, we continue our work on developing and operating the proposed methodologies in real time to index and share the obtained insights with the research and security communities at large, to strongly support and facilitate IoT security research using empirical data and aid in the IoT remediation objective at scale.

Acknowledgments

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.

References

[1] L. Atzori, A. Iera, and G. Morabito, "From 'Smart Objects' to 'Social Objects': The Next Evolutionary Step of the Internet of Things," *IEEE Commun. Mag.*, vol. 52, no. 1, Jan. 2014, pp. 97–105.
[2] "Hajime IoT Worm Infects Devices to Head Off Mirai; https://www.helpnetsecurity.com/2017/04/19/hajime-iot-worm/.
[3] T. Yu *et al.*, "Handling a Trillion (Unfixable) Flaws on a Billion Devices: Rethinking Network Security for the Internet-Of-Things," *Proc. 14th ACM Wksp. on Topics in Networks*, 2015, p. 5.
[4] Y. Ja. Jia *et al.*, "Contexiot: Towards Providing Contextual Integrity to Appified IoT Platforms," *Proc. 21st Network and Distributed System Security Symp.* , 2017.
[5] E. Fernandes *et al.*, "Flowfence: Practical Data Protection for Emerging Iot Application Frameworks," *Proc. USENIX Security Symp.*, 2016.

[6] B. Ur, J. Jung, and S. Schechter, "The Current State of Access Control for Smart Devices in Homes," *Proc. Wksp. Home Usable Privacy and Security*, 2013.
[7] E. Ronen and A. Shamir, "Extended Functionality Attacks on IoT Devices: The Case of Smart Lights," *Proc. 2016 IEEE Euro. Symp. Security and Privacy*, 2016, pp. 3–12.
[8] G. Ho *et al.*, "Smart Locks: Lessons for Securing Commodity Internet Of Things Devices," *Proc. 11th ACM on Asia Conf. Computer and Commun. Security*, 2016, pp. 461–72.
[9] A. Cui and S. J Stolfo, "A Quantitative Analysis of the Insecurity of Embedded Network Devices: Results of a Wide-Area Scan," *Proc. 26th Annual ACM Computer Security Applications Conf.*, 2010, pp. 97–106.
[10] A. Costin *et al.*, "A Large-Scale Analysis of the Security of Embedded Firmwares," *USENIX Security*, 2014, pp. 95–110.
[11] C. Fachkha *et al.*, "Internet-Scale Probing of CPS: Inference, Characterization and Orchestration Analysis," *Proc. NDSS*, vol. 17, 2017.
[12] R. Bodenheim *et al.*, "Evaluation of the Ability of the Shodan Search Engine to Identify Internet-Facing Industrial Control Devices," *Proc. Int'l. J. Critical Infrastructure Protection*, vol. 7, no. 2, 2014, pp. 114–23.
[13] C. Fachkha and M. Debbabi, "Darknet as a Source of Cyber Intelligence: Survey, Taxonomy, and Characterization," *IEEE Commun. Surveys & Tutorials*, vol. 18, no. 2, 2016, pp. 1197–1227.
[14] The CAIDA UCSD Near-Real-Time Network Telescope: Dec. 2016–Feb. 2017.
[15] L. Chi and X. Zhu, "Hashing Techniques: A Survey and Taxonomy," *ACM Computing Surveys*, vol. 50, no. 1, 2017, p. 11.

Biographies

Farooq Shaikh is pursuing his Ph.D. degree in electrical engineering at the University of South Florida (USF), where he is also President of the USF Whitehatters Computer Security Club (ethical hackers). He received his masters degree in electrical engineering from USF in 2017 and his bachelors degree in electronics engineering from Mumbai University (India) in 2015. Prior to joining USF he worked as a CCNA and CCNP certification exam trainer and also completed internships at Siemens Ltd and Jet Airways. His current research interests include IoT and cyberphysical systems security, IoT malware analysis, denial of service (DoS) attack mitigation techniques, deep learning/artificial intelligence, and security applications for SDN technologies. He is also actively involved in a wide range of cybersecurity training and outreach initiatives with local high-schools and organizations.

Elias Bou-Harb is an assistant professor in the Computer Science Department at Florida Atlantic University, where he directs the Cyber Threat Intelligence Laboratory. Previously, he was a visiting research scientist at Carnegie Mellon University. He is also a research scientist at the National Cyber Forensic and Training Alliance (NCFTA) of Canada. Elias holds a Ph.D. degree in computer science from Concordia University, Montreal, Canada. His research and development activities and interests focus on the broad area of operational cyber security, including, attacks detection and characterization, Internet measurements, cyber security for critical infrastructure and big data analytics.

Nataliia Neshenko is a Ph.D. student at Florida Atlantic University (FAU). She received her M.S. degree in computer science from FAU in 2018, an M.S. degree in management of organization from Kyiv Institute of Investment Management, Ukraine in 2004, and an M.S. degree in applied mathematics from Dnipro State University, Ukraine in 2000. She has over 16 years of project management experience, and holds PMP, PMI-ACP, and FCCA certifications. Her current research interests are in the areas of operational cybersecurity, including attacks detection and characterization, risk assessment methodologies, Internet of Things and data analytics.

Andrea P. Wright is a McKnight Doctoral Fellow and Sloan Scholar majoring in electrical engineering at the University of South Florida while serving as a Lieutenant in the Navy Reserves Information Warfare Community. She completed her Bachelor's degree in computer science at Bucknell University in 2008 and her Master's degree in computer engineering at the University of New Mexico. She has served as a data security analyst and an information systems security engineer (ISSE) at Booz Allen Hamilton. She also worked for Textron Systems (formerly AAI) as a level III information assurance engineer. Her current research focuses on securing software defined networks (SDN) against large-scale denial of service (DoS) attacks.

Nasir Ghani is a professor of electrical engineering at the University of South Florida and research liaison for Cyber Florida, a state-funded center focusing on cybersecurity research, education, and outreach. Prior to this he was the Associate Chair of the Electrical and Computer Engineering Department at the University of New Mexico (2007-2013) and a faculty member at Tennessee Tech University (2003-2007). He has also spent several years working in industry at large blue chip organizations (IBM, Motorola, Nokia) and several hi-tech startups. His research interests include cyberinfrastructure networks, cybersecurity, cloud computing, disaster recovery, and IoT/cyberphysical systems. He has published over 200 peer-reviewed articles and has several highly-cited U.S. patents. In the past he has served as an associate editor for the *IEEE/OSA Journal of Optical and Communications and Networking*, *IEEE Systems*, and *IEEE Communications Letters*. He has also guest-edited special issues of *IEEE Network* and *IEEE Communications Magazine* and has chaired/organized symposia for numerous flagship IEEE conferences (such as IEEE Globecom, IEEE ICC, and IEEE ICCCN). He was also the Chair of the IEEE Technical Committee on High Speed Networking (TCHSN) from 2007-2010. He received his Ph.D. in electrical engineering from the University of Waterloo, Canada.