A Large-scale Empirical Study on the Vulnerability of Deployed IoT Devices

Binbin Zhao, Shouling Ji, Wei-Han Lee, Changting Lin, Haiqin Weng, Jingzheng Wu, Pan Zhou, Liming Fang, Raheem Beyah

Abstract—The Internet of Things (IoT) has become ubiquitous and greatly affected peoples' daily lives. With the increasing development of IoT devices, the corresponding security issues are becoming more and more challenging. Such a severe security situation raises the following questions that need urgent attention: What are the primary security threats that IoT devices face currently? How do vendors and users deal with these threats?

In this paper, we aim to answer these critical questions through a large-scale systematic study. Specifically, we perform a ten-month-long empirical study on the vulnerability of $1,362,906$ IoT devices varying from six types. The results show sufficient evidence that N-days vulnerability is seriously endangering the IoT devices: $385,060$ (28.25%) devices suffer from at least one N-days vulnerability. Moreover, $2,669$ of these vulnerable devices may have been compromised by botnets. We further reveal the massive differences among five popular IoT search engines: *Shodan* [1], *Censys* [2], [3], *Zoomeye* [4], *Fofa* [5] and *NTI* [6]. To study whether vendors and users adopt defenses against the threats, we measure the security of MQTT [7] servers, and identify that $12,740$ (88%) MQTT servers have no password protection. Our analysis can serve as an important guideline for investigating the security of IoT devices, as well as advancing the development of a more secure environment for IoT systems.

Index Terms—IoT Search Engine, Vulnerable Device Assessment.

1 Introduction

The Internet of Things (IoT) has become an essential part of Internet connectivity and offered great convenience to our daily lives. For instance, router changes the way of surfing the Internet; smart door lock avoids the cumbersome process of door opening; voice assistant (e.g., Amazon Echo and Google Home) enables us to interact with Internet services and other smart devices through voice commands. According to Gartner, there will be more than 20 billion IoT devices all over the world in 2020 [8]. Meanwhile, the booming of IoT devices also raises public's concern about their security risks and several real-world attacks further aggravate this panic. For instance, Mirai infected millions of IoT devices including IP cameras, DVRs, and routers, to form a botnet and launch DDoS attacks against various online services [9]. Also, hackers can use light to compromise several voice controlled IoT devices, such as smart speakers [10].

The vendors of IoT devices try to mitigate the security threat by making the source code of firmware unavailable to the public, e.g., through obfuscating the source code of firmware or disabling users' access to the firmware. They mainly defend against IoT attacks from the perspective of software implementation, which, nevertheless, neglects the a prominent problem - *N-days vulnerability attack*. A typical example is that Mirai adopts this basic attack to control millions of IoT devices [11]. In order to avoid this attack, vendors can require users to provide automatic firmware update mechanisms. However, with limited security awareness, most vendors do not actively offer this essential defense in their products. Moreover, due to the misconfiguration, a significant number of IoT devices are exposed to the public Internet, which makes this attack even serious. Therefore, the security issues of IoT devices are still one of the most challenging problems in the progress of IoT development. Towards having an in-depth understanding, in this paper, we systematically evaluate the security of IoT devices to figure out (i) the comprehensive view of understanding the current security status of IoT devices; (ii) the primary challenges they are facing currently; and (iii)

- *B. Zhao is with the College of Computer Science and Technology at Zhejiang University, Hangzhou, Zhejiang, 310027, China, and also with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332. E-mail: binbin.zhao@gatech.edu.*
- *S. Ji is with the College of Computer Science and Technology at Zhejiang University, Hangzhou, Zhejiang, 310027, China, and also with the Zhejiang University NGICS Platform, Hangzhou, 310027, China. Email: sji@zju.edu.cn.*
- *W. Lee is with the IBM T.J. Watson Research Center. Email: wei-han.lee1@ibm.com.*
- *C. Lin is with the Zhejiang Gongshang University, and the State Key Laboratory of Information Security (Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093). Email: linchangting@gmail.com.*
- *H. Weng is with the Ant Group, Hangzhou, China. Email: haiqin.wenghaiqin@antgroup.com.*
- *J. Wu is with the Institute of Software, ChineseAcademy of Sciences, Beijing. Email: jingzheng08@iscas.ac.cn.*
- *P. Zhou is with the Hubei Engineering Research Center on Big Data Security, School of Cyber Science and Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China. Email: zhoupannewton@gmail.com.*
- *L. Fang is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics. Email: fanglming@nuaa.edu.cn.*
- *R. Beyah is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332. Email: raheem.beyah@ece.gatech.edu.*
- ** This work was partially conducted when B. Zhao was at Zhejiang University.*

how vendors and users deal with these challenges.

We take the first step toward evaluating the vulnerabilities of IoT devices by using five popular and powerful IoT search engines, namely *Shodan* [1], *Censys* [2], [3], *Zoomeye* [4], *Fofa* [5] and *NTI* [6]. More specifically, we evaluate the security vulnerability of IoT devices with 73 known N-days vulnerabilities. We then discuss the relationship between the vulnerability results of IoT devices with their locations. To this end, we collect $8,554,183$ IoT devices and finally obtain $1,362,906$ potentially vulnerable IoT devices as our dataset (after data preprocessing and the corresponding details will be discussed in Section 3.3), which vary from six types and involve 24 vendors. We also explore the geographical difference in the security of IoT devices. Besides, we test the no-password protection problem on $14,477$ MQTT servers that deploy on Amazon Web Services, Alibaba Cloud, Google Cloud, Microsoft Azure and Tencent Cloud.

Contributions. In summary, this paper mainly makes the following contributions:

- To the best of our knowledge, we conduct the first systematic study to evaluate five existing popular IoT search engines: *Shodan*, *Censys*, *Zoomeye*, *Fofa* and *NTI*. We reveal the significant differences among them regarding search ability, data accuracy rate, responding time and scanning period based on which we figure out the suitable application scopes for each IoT search engine, providing useful guidelines for IoT devices collection.
- We conduct thus far the largest empirical study on the vulnerability of $1,362,906$ IoT devices and $14,477$ MQTT servers. We have identified that $385,060$ (28.25%) IoT devices are vulnerable to the N-days vulnerability attack, and $12,740$ (88%) MQTT servers have no password protection. We further reveal that $2,669$ devices may have already been infected by botnets. Besides, we confirm the geographical difference in the security of IoT devices that most vulnerable devices are mainly located on few countries, e.g., U.S. and China.

2 Background

In this section, we first introduce five IoT search engines that we compare in our work. We then discuss four IoT security challenges that require serious consideration.

2.1 IoT Search Engines

In 2005, the International Telecommunication Union (ITU), which is responsible for issues that concern information and communication technologies, published a summary indicating that the age of IoT was coming [12]. Nevertheless, with the increasing development of IoT devices, more and more security related problems also emerged. The general-purpose search engines, such as Google and Bing, are not efficient in searching IoT devices. The existing pull models of information exchange, where the web search engines use web crawlers to discover web server information, do not work for most IoT cases. In 2009, the world's first IoT search engine *Shodan* [1] was brought online, which was designed to search the Internet-connected devices. Then, a variety of other search engines have been designed such as *Censys* [2], [3], *Zoomeye* [4], *Fofa* [5] and *NTI* [6]. Security researchers as well as attackers can use these IoT search engines to understand the component coverage and the damage scope of vulnerabilities of IoT devices. Considering that our work dedicates to evaluating the security of IoT devices at a large-scale, it is essential for us to choose appropriate search engines to collect data. In this work, we mainly focus on five IoT search engines described in detail as follows.

Shodan is the world's first search engine for Internet-connected devices [1]. *Shodan* is used around the world by large enterprises and security researchers. It deploys multiple servers located all over the world, and these servers provide 24/7 continuous detection through the Internet. Due to the ongoing scanning, *Shodan* offers the latest Internet intelligence, and users can know the influence of a specific component or the vulnerability influence at the Internet-scale. *Shodan* also provides public APIs for users, which assist users in accessing all of *Shodan*'s data more conveniently.

Censys is a platform that helps information security practitioners discover and analyze IoT devices that are accessible from the Internet [2]. Zakir et al. [3] released the first version of *Censys* in 2015 and now a team includes the world's leading experts on Internet-wide security is supporting this search engine. *Censys* has performed thousands of Internet-wide scans over the past five years, consisting of trillions of probes which play an important role in discovering and analyzing several serious Internet-scale vulnerabilities: FREAK [13], Heartbleed [14], and Mirai [9].

Zoomeye is a cyberspace search engine from China [4], which is dedicated to recording information of devices, websites, services and components etc. *Zoomeye* has two powerful detection engines Xmap and Wmap targeting devices and websites in the cyberspace, respectively. Security researchers can use it to identify IoT devices through 24/7 continuous detection. *Zoomeye* is designed for threat detection and situational awareness at Internet-scale.

Fofa is a search engine for IoT devices, which aims to make real-world data accessible and actionable in a secure and privacy-preserving manner [5]. *Fofa* provides a vulnerability market for its VIP users, which contains thousands of scripts generated from PoC of N-days vulnerabilities. Users can buy these scripts and leverage them to evaluate the security of IoT devices.

NTI is a non-profit search engine yet not publicly opened and is supported by NSFOCUS now [6]. *NTI* has detected thousands of botnet Command & Control servers and millions of infected hosts in 2018. In addition, *NTI* conducts a survey about the IoT security status each year.

These five IoT search engines all have the ability in scanning the entire IPv4 public Internet within a short period of time. The marked differences among them are their different scanning periods, scanning strategies, the amount of IoT device banner information they collected, and the accuracy rates of their saved data. We illustrate the comparison results of these five IoT search engines in Section 3.2. Besides the above five IoT search engines, there exist other two IoT search engines, *Thingful* [15] and *IoT Crawler* [16]. *Thingful* provides the data mainly for the commercial purpose rather than the security study and *IoT Crawler* is still under development. Therefore, these two IoT

search engines are not considered in our test.

2.2 IoT Security Challenges

IoT devices are omnipresent in our everyday life nowadays and provide great convenience for users. However, they also bring serious security problems. Heer et al. [17] discussed the security challenges in the IP-based IoTs, where they primarily focused on the challenges of the communication among IoT devices. Different from their work, we present three security challenges that have increasingly raised for modern IoT devices.

Misconfiguration. With the help of IoT search engines, it is not difficult to collect millions of IoT devices that are connected to the Internet. For example, with the simple keyword *router*, people can obtain more than ten million related records from *Zoomeye* collected since 2018. The key reason for most of these exposed IoT devices is the misconfiguration of these devices' owners. Due to the limited security awareness or computer network knowledge, a significant number of users may open the wide area network (WAN) unknowingly or configure the NAT-DDNS without any protections. However, the situation is becoming more critical since the search engines can collect hundreds of millions of newly exposed IoT devices each year. For instance, *Zoomeye* has collected over one billion IoT devices since it has been established. Thus, misconfiguration is one of the most severe challenges that IoT devices face currently.

Default Credential. Most vendors offer the default account and password for their devices, which, however, may pose new threats to the security of IoT devices. It leaves the major source of hidden danger. For instance, Mirai, the most notorious botnet, leverages the default passwords to affect millions of IoT devices. In fact, default credential is a widespread security problem which is not limited to IoT devices. Though this security problem has existed for so many years, we have not seen any mitigating trends. Meanwhile, it is surprising to notice that vendors rarely offer countermeasures to mitigate this serious problem.

Vulnerability Attack. Zero-day vulnerability attacks and N-days vulnerability attacks are seriously endangering the security of all the IoT devices. It is unrealistic to solve the Zero-day vulnerability thoroughly since the developer cannot avoid the bug even if they are very cautious in programming and code audit. Vendors have adopted several strategies to mitigate this problem. For example, they deploy the honeypots all over the world which can catch the Zero-day vulnerabilities in the wild, and some vendors choose not to disclose their firmware since hackers can decompile the firmware to find the flaws. However, these countermeasures are very inefficient since the honeypots need to wait for the attacks passively and hackers can obtain the firmware by other methods. Besides, the traditional fuzzing method cannot directly apply to IoT firmware [18], [19], [20]. While public pays more attention to Zero-day vulnerabilities, N-days vulnerabilities actually bring even serious risk to IoT devices [21]. N-days vulnerabilities are a goldmine for hackers since the exploitation of these vulnerabilities are already publicly known. With the help of IoT search engines, hackers can attack the exposed devices through the Internet easily by using N-days vulnerabilities [22], [23], [24].

3 Evaluation Methodology and Setup

In this section, we first introduce the IoT devices that we focus on in this paper. Then, we illustrate our experiments on comparing the aforementioned five IoT search engines. Besides, we describe how we collect data from the IoT search engines and how we preprocess the raw data. Finally, we introduce our evaluation scope.

3.1 Device Selection

While there are various types of IoT devices exposed to the Internet, we focus on six representative types, of which the three general-purpose devices and the other three are specialized devices. For the general-purpose devices, we select *router*, *IP camera*, and *printer* since they have been broadly utilized in our daily lives. Previous works pay much attention to the security of these three kinds of devices [22], [25], [26]. For the specialized devices, we select *mining device*, *medical device*, and *Industrial Control System (ICS)*. Currently, there are few works that focus on their security, especially for the mining devices and medical devices. Nevertheless, they are very important sectors in IoT. Therefore, it is necessary and meaningful to conduct a comprehensive study to understand their security. In addition to these six types, our analysis can also be generally extended to other types of IoT devices in a straightforward manner.

Router is one of the mostly widespread IoT devices and therefore, it is also the hacker's primary target. Even worse, the large expose of routers to the Internet makes them more vulnerable to the hackers. In our work, we focus on routers and study their vulnerabilities. To make our research more representative and concentrated, we choose five popular types of routers, namely HUAWEI, TP-Link, D-Link, ASUS, and MikroTik.

IP camera is indispensable in our daily life and has millions of end users. It is reported that due to the misconfiguration or other possible reasons, at least millions of IP cameras are exposed to the Internet. Of course, these exposed IP cameras become the ideal targets of the botnets. In this paper, we choose some of the most popular IP camera vendors including Hikvision, DAHUA, Axis, Avtech, and Netwave, to explore the security threat they are facing currently.

Printer is an essential part of our life and mostly used in the social organizations, e.g., companies and colleges. However, the vendors and the users have limited security awareness of printers, and according to the Spiceworks Survey [27] conducted by HP, only 16% of respondents think that printers are under the high security risk. In our work, we choose to analyze some of the most popular printer vendors, including HP, Brother, EPSON, Canon, and SAMSUNG. We use these devices as the representative examples to evaluate the current security status of printers.

Mining Device, as a novel IoT device, has been rapidly developed due to the increasing popularity of cryptocurrency, such as Bitcoin and Ethereum. Due to the relative high economic value, it has become the hacker's potential target. The attacks against mining devices have already been reported for several times currently [28]. Thus, we want to figure out the current security status of mining devices. Specially, we choose the Antminer [29] and Claymore [30] as the

target devices since the Antminer is the most well-known Bitcoin mining device and Claymore is widely adopted in mining the ethereum.

Medical Device is a relatively less popular IoT device that the public may easily neglect its security problem. These medical devices, which are closely related to the patients' privacy and health, however are exposed to the Internet, making them be the easy targets for the cyber attackers. In our work, we aim to figure out the current security status of the medical devices, and experimentally measure the vulnerabilities of GE Healthcare Softneta and Dicoogle.

Industrial Control System (ICS) is a more specialized Cyber-Physical System (CPS) that involves a great number of infrastructures. ICS devices usually run or operate a variety of customized protocols, e.g., Modbus and Siemens S7, in the application layer. However, by exploiting these customized protocols, we can easily identify the ICS devices exposed to the Internet, which brings potential security threats. Moreover, there already have reports of attacks on ICS. For instance, in 2010, the Stuxnet [31] caused substantial damage to Iran's nuclear program. The security of ICS is more critical than other IoT devices in a sense that it may endanger the national security. In our research, we consider Rockwell Automation, WAGO and Schneider Electric as the primary targets. All of these devices have multiple N-days vulnerabilities that are disclosed in CVE [32].

3.2 Search Engine Selection

As introduced in Section 2.1, there exist multiple IoT search engines. Their searching ability and data accuracy rate however vary greatly owing to the different searching techniques and data maintaining strategies they use. Since our measurement is based on the massive data collected from the IoT search engines, it is essential for us to select appropriate ones to collect data. Specifically, we evaluate five different IoT search engines *Shodan*, *Censys*, *Zoomeye*, *Fofa* and *NTI* with respect to four aspects, searching ability, raw data accuracy, responding time, and scanning period.

Searching Ability. We prefer to select such an IoT search engine that can provide sufficient data about all the tested IoT devices. Thus, we perform a preliminary comparison of the search engines in searching for specific devices. We select six representative devices, including TP-Link's router, Hikvision's IP camera, HP's printer, Bitmain's mining device, Dicoogle's PACS and ModbusGW. TP-Link, Hikvision, HP, and Bitmain are all leading vendors in their field. Besides, Dicoogle is one of the top 10 free open source PACS projects [33] and has been studied by previous work [34]. Moreover, ModbusGW is an essential part of the IoT ecosystem that has been widely adopted in a great number of application scenarios, such as connecting PLCs in the SCADA network [35]. Of each search engine, we use it to collect data about all the six types of devices in a same week, as shown in Table 1. As the table shows, *Zoomeye* collects the most exposed devices almost in all the tested types, while Shodan and Censys collect the least. The huge difference can be ascribed to two reasons: (i) IoT search engines adopt different data storage strategies. *Shodan* and *Censys* only present the records they collected at the current year while *Zoomeye*, *Fofa* and *NTI* provide all historical records they collected. (ii) IoT search engines vary greatly in their scanning techniques. *Zoomeye*, *NTI* and *Fofa* have more advanced scanning techniques which can help them identify more IoT devices. For instance, *Zoomeye* adopts two powerful detection engines Xmap and Wmap, which perform better than Nmap [36] and Zmap [37].

TABLE 1
Ability in Searching Specific IoT Devices

Devices	Vendors	Zoomeye	Shodan	Censys	Fofa	NTI
Router	TP-Link	1,047,861	248,062	261,149	924,950	743,119
IP camera	Hikvision	3,607,552	130,836	124,238	427,935	2,749,631
Printer	HP	210,507	119,183	87,343	148,309	127,625
Antminer	Bitmain	1,158	344	306	2,619	792
PACS	Dicoogle	10	1	2	1	0
ModbusGW	Modbus	47	14	8	9	36

TABLE 2
Data Accuracy Rate

Vendors	Amount	Zoomeye	Shodan	Censys	Fofa	NTI
Router						
MikroTik	50K	84.09%	68.45%	77.32%	75.97%	81.48%
TP-Link	50K	88.23%	81.05%	76.71%	84.33%	90.34%
IP camera						
Hikvision	50K	90.03%	79.53%	86.59%	83.29%	89.09%
Axis	50K	79.10%	74.68%	89.62%	71.83%	86.13%
Printer						
HP	10K	94.63%	91.35%	92.27%	90.64%	88.03%
Brother	10K	86.82%	87.13%	86.19%	93.77%	91.75%

Raw Data Accuracy. Second, we compare the accuracy rate of the raw data collected from IoT search engines. The accuracy rate is calculated as the ratio of valid data in the raw data. We perform this trial on six representative devices, including two routers from MikroTik and TP-Link, two IP cameras from Hikvision and Axis, and two printers from HP and Brother, as shown in Table 2. We use the method as discussed in Section 3.3 to collect raw data and, to ensure fairness, we collect all the device data over the same period of five months (from 01/01/2018 to 06/01/2018). For router and IP camera, we randomly collect 10,000 devices each month with a total of $50,000$ devices. Since *Censys* only collects a few exposed printers in this period, we randomly collect $2,000$ printers each month with a total of $10,000$ printers for two vendors individually.

We leverage the scanner, which will be discussed in Section 3.3 to filter out the invalid data in the raw data collected by each of the IoT search engines. As shown in Table 2, *Zoomeye* and *NTI* have better data accuracy than other IoT search engines for most cases. Therefore, *Zoomeye* and *NTI* are better candidates for users who prefer to use the raw data collected from IoT search engines.

Responding Time. Third, we analyze the responding time of each search engine to the newly exposed IoT devices. This is an important metric to assess each search engine's ability which can accurately reflect their sensitivity to the newly exposed devices. Towards this end, we deploy seven servers worldwide on 11/19/2018, along with SSH service at port 22 opened, which is widely used in a great number of

TABLE 3
Responding Time and Scanning Period

IP Address	Location	Creation Time	Zoomeye		Shodan		Censys		Fofa		NTI	
			Responding Time	Scanning Period	Responding Time	Scanning Period	Responding Time	Scanning Period	Responding Time	Scanning Period	Responding Time	Scanning Period
35.236.14.159	Los Angeles, U.S.	11/19/2018	12/07/2018	-	11/21/2018	16 days	11/20/2018	-	12/05/2018	19 days	11/27/2018	-
35.243.65.199	Tokyo, Japan	11/19/2018	12/07/2018	-	11/21/2018	16 days	11/20/2018	24 days	12/08/2018	22 days	12/05/2018	17 days
35.199.71.88	São Paulo, Brazil	11/19/2018	12/07/2018	-	12/04/2018	14 days	11/20/2018	-	12/10/2018	15 days	11/30/2018	-
47.88.63.131	Santa Clara, U.S.	11/19/2018	12/07/2018	-	12/10/2018	17 days	-	-	-	-	12/01/2018	18 days
47.75.43.192	Hongkong, China	11/19/2018	-	-	12/01/2018	15 days	12/11/2018	-	12/06/2018	16 days	12/02/2018	21 days
66.42.94.35	Atlanta, U.S.	11/19/2018	12/08/2018	19 days	11/22/2018	15 days	11/20/2018	-	12/07/2018	20 days	11/21/2018	-
45.77.232.219	Sydney, Australia	11/19/2018	12/14/2018	21 days	11/24/2018	17 days	11/20/2018	-	12/02/2018	18 days	12/06/2018	-

IoT devices. We keep these severs for 60 days. Three of the seven servers are from Google Cloud located in Los Angeles, Tokyo and So Paulo, two of them are from Alibaba Cloud located in Santa Clara and Hongkong, and the remaining two are from Vultr located in Atlanta and Sydney. We deploy seven servers to eliminate the accident error that the search engine has already scanned servers before. We query each IoT search engine per day, and obtain the returned records that include the latest indexed data of these servers. We label the date when the servers are scanned by search engines for the first time after deployment, which we regard as the responding time, as shown in Table 3. *Shodan* and *Censys* are more sensitive to the newly exposed servers than the other three IoT search engines. At the time when they are closed, the servers from the Alibaba Cloud are still not scanned by several search engines. To the best of our knowledge, Alibaba Cloud have designed countermeasures to protect against the scanning activities from IoT search engines, e.g., blocking the malicious IPs. From this perspective, *Shodan* and *NTI* are more efficient in scanning the servers or devices which have countermeasures to defend against the scanners.

Scanning Period. Finally, we measure the scanning period of IoT search engines. The scanning period is the time difference between two contiguous scannings. We list the scanning period for each IoT search engine in Table 3. From Table 3, we have the following important observations: (i) *Zoomeye*, *Censys* and *NTI* rarely scan the same server twice in a short time. We conjecture that they may ignore the server which has been scanned recently in order to save cost. (ii) *Shodan*, which scans the whole IPv4 space within every 17 days on average, has the shortest scanning period among these five search engines. (iii) *Shodan* is the only one that scans all the servers for more than once and *Fofa* scans the whole IPv4 space within nearly every 18 days on average.

From the evaluation results, we observe that each IoT search engine has its own advantages and disadvantages. We summarize the evaluation results and analysis into a set of recommendations for users and researchers. 1) We recommend the users who prefer to conduct their research on the most recent data to choose *Shodan* and *Fofa* since these IoT search engines scan the Internet more frequently. 2) *Shodan* and *Censys* are appropriate for users who want to collect the newly exposed IoT devices. 3) If the users prefer to collect large-scale data, *Zoomeye* and *NTI* are better candidates because these two IoT search engines store more historical data and have a higher data accuracy rate.

In our following experiments, we use the *Zoomeye* to collect routers, IP cameras, printers, medical devices and ICS devices, and adopt the *Fofa* to collect mining devices. Besides, we utilize *Shodan* to collect MQTT servers.

3.3 Data Collection and Preprocessing

Data Collection. With the help of IoT search engines, we can collect millions of IoT devices. However, it also requires extensive manual processing. In this paper, we first review the search syntax guides provided by *Zoomeye* and *Fofa* and construct $1,281$ keywords to query the IoT search engines. The keywords are mainly constructed based on two rules. The first rule is combining the vendor names with the device types and versions. For instance, we construct the keyword **router app:"TP-Link TL-WR841N"** to search the TL-WR841N router of TP-Link on *Zoomeye* and the keyword **app="Claymore-Miner"** to search the mining devices of Claymore on *Fofa*. The second rule is based on the network features of these relatively less popular IoT devices, e.g., ICS. For example, the Siemens S7 protocol always use the port 102 for communication. Thus, we can use the keyword **"port:102"** to search the IoT devices that use the Siemens S7 protocol on *Zoomeye*.

Next, we remove the keywords which will return unexpected results. We first query each constructed keyword on IoT search engines and analyze the returned results. If the returned results are different from our expectations, we regard the corresponding keywords as invalid and remove them from the database. For example, if a keyword which is initially designed for searching the mining devices of Claymore but it returns irrelevant results, we will delete this keyword. After we test these $1,281$ keywords, we finally remove 959 inappropriate keywords and obtain 322 valid keywords. More specifically, as shown in Table 4, we collect $4,278,226$ routers by using 104 keywords, $2,844,017$ IP cameras by using 82 keywords, $1,341,610$ printers by using 84 keywords, 60,025 mining devices by using 5 keywords, $1,897$ medical devices by using 25 keywords and $29,408$ ICS devices by using 22 keywords. With these 322 carefully selected keywords, we totally collect $8,554,183$ IoT devices in a week from $01/03/2019$ to $01/10/2019$ that vary from six categories and involve 24 vendors.

Data Preprocessing. Previous works neglect or fail to clearly explain how they deal with the invalid and outdated data [38]. Actually, the invalid and outdated data may seriously influence the performance of our measurement. Thus, we aim to preprocess the raw data by filtering out the invalid and outdated items. Specifically, our preprocessing consists of the following three steps.

First, we delete the devices which have the same IPs and remove the devices whose response time is over five seconds. For the devices with long response time, they may

TABLE 4
Data Selection and Preprocessing

	Total Data Variation			Vendors	Specific Data Variation			Keywords Variation		Records of Banner Information Variation	
	Raw Data	Processed	Final Selected		Raw Data	Processed	Final Selected	Initial Data	Final Selected	Initial Data	Final Selected
Router	4,278,226	1,349,076	500,000	HUAWEI	1,438,967	438,824	100,000	139	32	1,017	498
				TP-Link	984,259	416,327	100,000	116	31	862	219
				D-Link	1,031,242	247,098	100,000	69	17	426	153
				ASUS	514,624	138,241	100,000	76	9	916	411
				MikroTik	308,134	108,586	100,000	83	15	321	171
IP camera	2,844,017	973,520	500,000	Hikvision	802,371	341,471	100,000	91	23	852	309
				DAHUA	534,295	186,319	100,000	82	14	608	86
				Axis	413,287	121,072	100,000	36	16	240	72
				Avtech	630,112	213,928	100,000	49	16	192	39
				Netwave	463,952	110,730	100,000	40	13	201	15
Printer	1,341,610	402,721	310,000	HP	647,290	181,923	100,000	73	26	104	41
				Brother	483,124	105,343	100,000	39	21	63	19
				EPSON	86,478	51,368	50,000	58	20	117	53
				Canon	51,328	22,394	20,000	23	12	77	5
				Samsung	46,659	20,195	20,000	16	4	34	7
				Octoprint	26,731	21,498	20,000	1	1	1	1
Mining Device	60,025	39,121	38,000	Claymore	50,987	30,712	30,000	11	2	4	1
				Antminer	9,038	8,409	8,000	13	3	11	3
Medical Device	3,935	782	706	GE Healthcare	1,498	305	300	96	19	74	12
				Softneta	316	92	92	6	3	15	6
				Dicoogle	83	14	14	4	3	21	5
ICS	31,087	16,212	15,500	Rockwell Automation	9,731	5,183	5000	67	15	93	27
				Schneider Electric	18,874	9,364	9000	81	6	85	11
				WAGO	803	563	500	12	1	24	5

already stay away from the Internet or have a poor network condition which may significantly increase the test time and affect the credibility of the final measurement. After this step, $5,061,289$ devices are left.

Second, we collect sufficient banner information for each kind of device. Banner is widely used on IoT devices to display some important information, e.g., the information displayed when users login to the server by SSH, which can help identify the type of IoT devices. Nevertheless, with the increasing number of IoT devices, each kind of devices has its own banner information. Thus, in order to identify the IoT devices accurately, we need to collect the banner information as much as possible. We construct a list that contains the categories of IoT devices and corresponding vendors included in our raw dataset based on the constructed keywords. Then, we collect the banner information relates to the devices in the list. The IoT search engines listed in Section 2.1 provide millions of records of banner information they used for identifying devices. We first manually select $6,358$ related records of the banner information, where $5,622$ are collected from search engines and 736 are collected from the Internet and the user guide of the devices.

Third, we develop our scanner based on the fingerprinting technique. Fingerprinting is an effective method to identify the firmware version of IoT devices. Nmap [36] and Zmap [37] are two state-of-the-art fingerprinting tools that can utilize banner information to identify IoT devices. In our work, we choose Nmap rather than Zmap, since Nmap can provide much more details of target devices, and the corresponding scanning time is acceptable. First of all, the scanner will use Nmap to send HTTP requests to the target devices and receive their response data. The response data may contain the information of target devices, such as vendor name (e.g. TP-Link), device type (e.g. router), device model (e.g. TL-WR840N) and firmware version (e.g. 0.9.1). Thus, we then compare the response data with our collected banner information. If there is a match, we can confirm the above information of target devices. During our preliminary test, we find that some records of banner information are inappropriate. We remove $4,324$ improper records of banner information and $2,034$ records of banner information are left. Based on these valid records of banner information, the scanner supports identifying $1,516$ different models of 6 types of IoT devices from 24 vendors.

To test the accuracy of the scanner, we build a benchmark that has 1,000 already identified IoT devices which belong to 24 vendors we will test in this paper. For these 1,000 IoT devices, we manually analyze their response data and review their GUI interfaces to confirm their exact vendors, types, models and firmware versions. Next, we perform our scanner on the benchmark and finally achieves 97.2% accuracy. By using our scanner, we further remove $2,280,330$ invalid devices which can not be identified by fingerprints.

Through our data preprocessing, we totally detect $5,774,224$ invalid devices and obtain $2,779,959$ valid devices, including $1,349,076$ routers, $973,520$ IP cameras, $402,721$ printers, $39,121$ mining devices, 411 medical devices and $14,547$ ICS devices, as shown in Table 4. However, we do not perform the analysis on the entirety of the dataset due to three reasons. First of all, it is extremely time-consuming to analyze such a large dataset. Second, since the analysis will cost lots of time, some devices may already

change the IP address or stay away from the Internet during the experiment. Thus, we will keep a part of backup devices to replace these inactive devices. Finally, for statistical purposes and fairness, we try to adjust the number of the same type of devices from different vendors to be consistent. Thus, we randomly select 500K routers, 500K IP cameras, 310K printers, 38K mining devices, 406 medical devices, and 14.5K ICS devices, with a total of $1,362,906$ IoT devices in our evaluation dataset.

3.4 Evaluation Scope

In this paper, we mainly focus on the N-days vulnerability problem of the IoT devices. While public pays more attention to the zero-day vulnerability, N-days vulnerabilities actually bring more serious risks to IoT devices. With the help of IoT search engines, hackers can attack the exposed devices through the Internet easily by using N-days vulnerabilities. The PoC-Checking method is the most straightforward scheme to confirm the existence of N-days vulnerabilities [39]. However, it is illegal to test online IoT devices without ownership. The PoC-Checking method will trigger and exploit the vulnerability of real-world IoT devices, which may raise serious ethical concerns. Thus, we choose to leverage the firmware fingerprinting method to check whether the target devices are vulnerable [40]. The firmware fingerprinting technique will first send HTTP requests to the target devices and obtain their response data. It will then compare the response data from IoT devices with our collected banner information to identify their exact vendors, types, models and firmware versions. It supports identifying 97.2% devices in our dataset that across from 407 different models of 6 types of IoT devices from 24 vendors. Since IoT vulnerabilities have a strong connection with the firmware versions, we can check firmware versions of the target devices to determine its vulnerability. In our method, we first collect 73 N-days vulnerabilities, which have disclosed the affected firmware versions. Then we use our scanner, developed by the fingerprinting technique, to identify the firmware versions of the $1,362,906$ IoT devices from six categories. Our objective is to figure out the proportion of vulnerable devices and reveal the severity of existing IoT devices regarding N-days vulnerabilities.

4 Results and Analysis

In this section, we first analyze the unpatched N-days vulnerability problem across millions of IoT devices.Then, we conduct a further evaluation to explore the trending of the vulnerable rate of IoT devices for six months. The evaluation results reveal several flaws of the existing countermeasures of vendors in preventing the unpatched N-days vulnerability. Finally, we explore the relationship between the vulnerability results of IoT devices and their locations.

4.1 Vulnerability Evaluation

We first collect 73 N-days vulnerabilities from the CVE [32], EXPLOIT DATABASE [41] and SEEBUG [42] which have disclosed the affected firmware versions, as shown in Table 5. Then, we use our scanner, developed based on the fingerprint technique, to identify the firmware versions of target devices accurately. For the IoT devices, which are still using the affected firmware versions, we regard them as vulnerable devices.

Router Evaluation. First, we select 33 N-days vulnerabilities to evaluate the security of routers from five vendors, HUAWEI, TP-Link, D-Link, ASUS and MikroTik, with 100,000 routers per vendor and the corresponding results listed in Table 5. From Table 5, we notice that $31,363$ routers suffer from at least one N-days vulnerability. MikroTik has the highest vulnerable rate among these five vendors. 18.04% MikroTik routers suffer from a newly disclosed vulnerability CVE-2018-14847 which was disclosed one month before our test. For ASUS, we find 4.03% routers are vulnerable to a five-year-long N-days vulnerability, CVE-2014-9583, which alerts us that we cannot ignore the severity of the aged vulnerability even though it has been disclosed for many years. HUAWEI, TP-Link and D-Link have a relatively low vulnerable rate (approximately 2%). Through our extra study, we find that these three vendors all provide automatic update mechanisms for the latest routers which may result in their low vulnerable rate.

IP camera Evaluation. Second, we show our evaluation on IP cameras from five popular vendors with 100,000 IP cameras from each vendor. In total, we employ 18 vulnerabilities with the detailed results listed in Table 5. We observe that the IP camera has a much more serious security status than that of the router, where the vulnerable rates for all the involved vendors are higher than 10%. Netwave reaches an astonishing vulnerable rate that 80.42% IP cameras are vulnerable to at least one N-days vulnerability. All of these four vulnerabilities, that we selected to test the IP cameras of Netwave, have been disclosed at least three months before our test. Thus, we assume that Netwave does not push the updates to the vulnerable devices timely. We select four vulnerabilities to test Axis and Avtech, where both of them achieve high vulnerable rates with 35.14% and 46.57%, respectively. Hikvision and DAHUA have relatively low vulnerable rates among these five vendors. It is worthy noting that DAHUA suffers greatly from a six-year-long vulnerability, CVE-2013-6117. In addition, since these two vendors have exposed the most IP cameras to the Internet, reaching tens of millions of scale, their security status are much more serious than the other three vendors.

Printer Evaluation. Third, we evaluate the security of printers from six well-known vendors. Though HP has provided automatic update mechanisms for its printers, it still has a high vulnerable rate of 25.52%. Possible reasons are that 1) users need to manually activate the automatic update mechanisms and 2) the mechanisms only exist in the most recent versions of printers. Canon has a relatively low vulnerable rate among these vendors, which also reaches 16.99%, and a six-year-long vulnerability still affects 4.20% tested printers. Brother and EPSON have extremely high vulnerable rates where more than 50% of the tested printers are vulnerable. Compared to HP, these two vendors do not provide automatic update mechanisms for their printers. We only find one appropriate vulnerability, CVE-2012-4964, to test SAMSUNG printers, which has been disclosed for seven years. We observe that there are still 73 SAMSUNG printers that suffer from this vulnerability. Octoprint provides a web interface for 3D printers that can control and monitor the

TABLE 5
Large-scale Vulnerability Evaluation

	Vendor	Dataset	Total Vulnerable Devices		N-days Vulnerabilities	Severity	Vulnerable Devices	
			Amount	Vulnerable Rate			Amount	Vulnerable Rate
Router	HUAWEI	100K	885	0.885%	CVE-2017-17215	High	601	0.60%
					CVE-2015-7254	Medium	285	0.29%
	TP-Link	100K	2,193	2.19%	CVE-2018-17004_17018	Medium	19	0.02%
					CVE-2017-16957	High	874	0.87%
					CVE-2018-11714	Critical	1,309	1.31%
	D-Link	100K	1,472	1.47%	CNVD-2018-01084	High	718	0.72%
					CVE-2018-9032	Critical	458	0.46%
					CVE-2017-9675	High	192	0.19%
					CVE-2018-10106	Critical	107	0.11%
	ASUS	100K	8,216	8.22%	CVE-2014-9583	Critical	4,027	4.03%
					CVE-2017-14698	Critical	2,410	2.41%
					CVE-2017-5891	High	936	0.94%
					CVE-2017-5892	High	955	0.96%
	MikroTik	100K	18,597	18.60%	CVE-2018-14847	Critical	18,036	18.04%
					CVE-2018-10070	High	317	0.32%
					CVE-2018-1156_1159	High	647	0.65%
IP camera	Hikvision	100K	18,104	18.10%	CVE-2014-4878	Critical	2,982	2.98%
					CVE-2014-4879	Critical	3,147	3.15%
					CVE-2014-4880	Critical	3,970	3.97%
					CVE-2017-7923	High	9,261	9.26%
	DAHUA	100K	13,499	13.50%	CVE-2013-6117	High	11,815	11.82%
					CVE-2017-7253	High	1,726	1.73%
	Axis	100K	35,137	35.14%	CVE-2018-9158	High	10,883	10.88%
					CVE-2018-10660_10662	Critical	24,711	24.71%
	Avtech	100K	46,571	46.57%	SSV-97347	High	14,950	14.95%
					SSV-97159	Medium	2,945	2.95%
					SSV-92493	High	5,606	5.61%
					SSV-92494	High	27,352	27.35%
	Netwave	100K	80,416	80.42%	CVE-2018-6479	High	15,081	15.08%
					CVE-2018-11653	Critical	40,138	40.14%
					CVE-2018-11654	High	68,030	68.03%
					CVE-2018-17240	High	38,019	38.02%
Printer	HP	100K	25,518	25.52%	CVE-2017-2741	Critical	25,518	25.52%
	Canon	20K	3,398	16.99%	CVE-2018-11692	Critical	2,559	12.80%
					CVE-2013-4615	Medium	839	4.20%
	Brother	100K	65,938	65.94%	CVE-2017-7588	Critical	64,407	64.41%
					CVE-2018-11581	Medium	3,206	3.21%
	EPSON	50K	43,582	87.16%	CVE-2018-5550	Medium	42,903	85.81%
					CVE-2018-14899_14900	Medium	1,882	3.76%
	SAMSUNG	20K	73	0.37%	CVE-2012-4964	High	73	0.37%
	Octoprint	20K	20,000	100%	CVE-2018-16710	Critical	20,000	100%
Mining Device	Claymore	30K	1,034	3.45%	CVE-2018-6317	Critical	382	1.27%
					CVE-2018-1000049	High	72	0.24%
					CVE-2017-16930	Critical	580	1.93%
	Antminer	8K	4	0.04%	CVE-2018-11220	High	4	0.04%
Medical Device	GE Healthcare	300	2	0.67%	CVE-2017-14002	Critical	1	0.33%
					CVE-2017-14006	Critical	1	0.33%
					CVE-2017-14008	Critical	0	0
	Softneta	92	0	0	EDB-ID-45347	Medium	0	0
	Dicoogle	14	0	0	EDB-ID-45007	Critical	0	0
ICS	Rockwell Automation	5K	12	0.24%	CVE-2018-19616	High	12	0.24%
	WAGO	500	2	0.4%	CVE-2018-16210	Medium	2	0.4%
	Schneider Electric	9K	407	4.52%	CVE-2017-6026	Critical	407	4.52%

printers' activities. Nevertheless, all the tested Octoprint printers are vulnerable to CVE-2018-16710 which allows unauthenticated users to download the project files from the printers. In addition, Octoprint does not regard it as a vulnerability and only ascribes it to the users' limited security awareness.

Mining Device Evaluation. Then, we evaluate the security of mining devices produced by two vendors, Claymore and Antminer, with four N-days vulnerabilities. We find that 3.45% Claymore mining devices suffer from at least one N-days vulnerability. For Antminer, we discover that only 4 devices can be attacked by CVE-2018-11220. Through our further study, we find the Antminer will push the firmware update notification to users timely when the vulnerability is repaired, which can help to explain such a low vulnerable rate. With more and more mining devices exposed to the Internet, the N-days vulnerability attack problem should be seriously considered by mining devices vendors.

Medical Device Evaluation. Furthermore, we investigate the security of medical devices from three vendors, GE Healthcare, Softneta and Dicoogle. According to our results, medical devices suffer slightly from the N-days vulnerability, where only two unsecured devices of GE Healthcare are identified. However, we cannot neglect the potential security threat since each unsecured medical device may bring massive loss, even endanger the lives of patients.

ICS Evaluation. Finally, we choose three N-days

TABLE 6
Further Vulnerability Evaluation

N-days Vulnerability	Vendors	Vulnerable Version	Disclosure Date
Router			
CVE-2017-17215	HUAWEI	HG532	12/04/2017
CVE-2018-11714	TP-Link	TL-WR840N/841N	06/04/2018
CVE-2018-9032	D-Link	DIR-850L	03/26/2018
CVE-2018-6000	ASUS	Before 3.0.0.4.384_10007	01/22/2018
CVE-2018-14847	MikroTik	RouterOS before 6.43rc3	08/02/2018
IP camera			
CVE-2017-7923	Hikvision	V5.2.0-V5.4.0	04/18/2017
CVE-2018-10661	Axis	Multiple models	05/02/2018
SSV-92494	Avtech	All firmware versions	10/25/2016
CVE-2018-11654	Netwave	All IP cameras	06/01/2018
Printer			
CVE-2017-2741	HP	PageWide/OfficeJet Printers	12/01/2016
CVE-2018-11692	Canon	LBP6650/3370/3460/7750C	06/04/2018
CVE-2017-7588	Brother	Multiple models	04/08/2017
CVE-2018-5550	EPSON	Epson AirPrint	01/12/2018
CVE-2018-16710	Octroprint	OctoPrint through 1.3.9	09/07/2018

vulnerabilities to measure the security of ICS from three vendors, Rockwell Automation, WAGO and Schneider Electric. Schneider Electric has the highest vulnerable rate among these ICS vendors, which reaches 4.52%. Rockwell Automation and WAGO have very low vulnerable rates (below 0.5%). However, the vulnerable ICS devices may bring much more serious security issues than the general-purpose IoT devices, e.g., router and IP camera. Therefore, the security of ICS regarding N-days vulnerabilities still requires serious consideration.

In addition to the above findings, we also notice that vulnerable rate varies greatly from different kinds of IoT devices. The vulnerable rates of IP camera and printer are significantly higher than other devices. Most of the unsecured devices are general-purpose IoT devices, e.g., router, IP camera and printer. The specialized devices have a very low vulnerable rate in comparison with the general-purpose IoT devices. However, we should pay more attention to the security of specialized devices since they may bring more massive loss than unsecured general-purpose devices.

In summary, we have identified that 385,060 devices still suffer from at least one N-days vulnerability.

4.2 Further Vulnerability Evaluation on Specific Devices

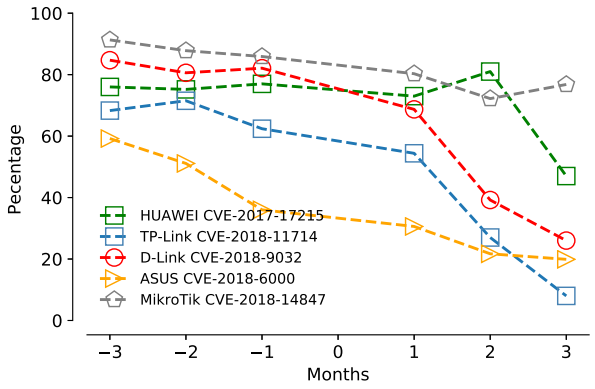
According to our large-scale vulnerability evaluation in Section 4.1, we find that the routers, IP cameras and printers suffer greatly from the N-days vulnerabilities. In this section, we explore the trending of the vulnerable rate of these three kinds of devices for six months, which can help us evaluate the vendors' responses to the N-days vulnerabilities. Specifically, we perform our evaluation on routers, IP cameras and printers from 14 popular vendors with one specific N-days vulnerability per vendor, as shown in Table 6. We collect the historical data of six months for each device with the range from three months before and three months after the vulnerabilities were disclosed. We only select the devices whose corresponding versions are influenced by vulnerabilities.

First, we evaluate the vulnerable rate trending on routers from five vendors with 1,000 routers per vendor per month, as shown in Figure 1(a). MikroTik's vulnerable rate varies little among six months even though the MikroTik has released the patched firmware on /04/23/2018 which is four months before the vulnerability was released. A possible reason is that MikroTik does not notify all users of vulnerable devices. In addition, the complicated firmware updating process of MikroTik may also be an important factor to the invariant vulnerable rate. Similar trends can also be observed for the other four vendors. The routers that are exposed to the Internet before the vulnerabilities were disclosed still have high vulnerable rate while in comparison the devices that are exposed after the disclosure time have low vulnerable rate. We infer that the reason for the downward trend is that the IP cameras with patched firmware have entered the market. Even though these vendors have already published the patched firmware (e.g., HUAWEI published the patched firmware of CVE-2017-17215 on 02/06/2018), most users have not updated the firmware yet. We conclude two possible reasons for this serious security status: (i) tedious manual updating process troubles a great number of users; (ii) vendors do not have an effective notification method to notify all vulnerable users.

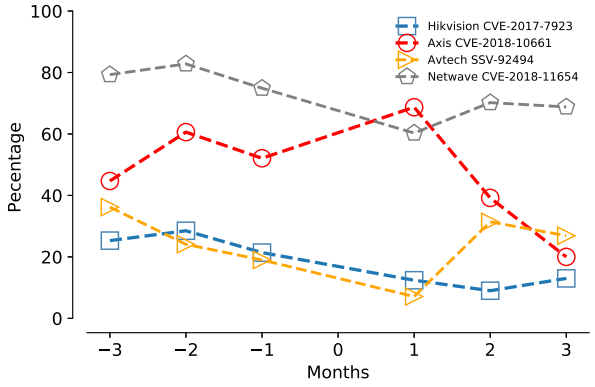
Second, we leverage four vulnerabilities to figure out the vulnerable rate of IP cameras from four vendors with 1,000 IP cameras per vendor per month, as shown in Figure 1(b). Avtech and Hikvision have steady low vulnerable rates for six months, which indicates that they have greatly controlled the vulnerabilities. Netwave has an extremely high vulnerable rate for six months with a slight downward trend. Axis's vulnerable rate decreases dramatically since one month after the vulnerability was disclosed. A possible reason is that the newly produced Axis IP cameras with patched firmware have entered the market at that time.

Finally, we show our evaluation of five vendors' printers by using five specific vulnerabilities with 500 printers per vendor per month. Note that we do not conduct experiments on SAMSUNG since no IoT search engine provides any record of SAMSUNG printer before 2014. As shown in Figure 1(c), HP has a relatively steady low vulnerable rate for six months. We ascribe this low rate to its automatic update mechanisms. Nevertheless, around 25% HP printers have not patched the vulnerability, CVE-2017-2741, yet. For these vulnerable printers, we notice that they still use outdated firmware. We infer that some of them may close the automatic update mechanisms or HP does not provide automatic update mechanisms on these vulnerable devices. Canon has a higher vulnerable rate than HP which is still lower than the other three vendors. EPSON and Brother both have a high vulnerable rate for six months without any downward trend. Though both of them allow users to manually update the printers, the vulnerability to N-days attack is not mitigated. Furthermore, the manually updating process is complicated where the users need to download the firmware from the official website and upload it to the printers through the embedded web interface. Octoprint maintains 100% vulnerable rate for six months, which indicates that the vendor does not develop any countermeasures to mitigate this severe vulnerability.

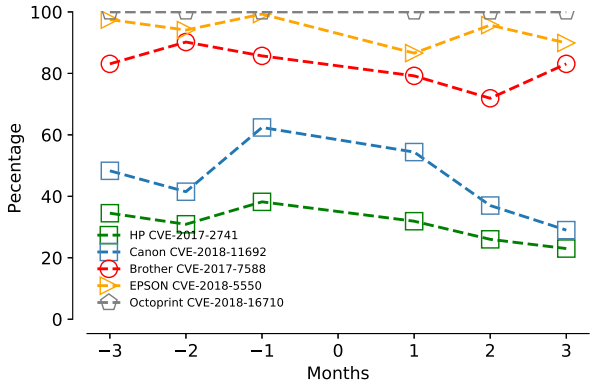
According to our further vulnerability evaluation, we find that some vendors do not have an effective notification method to inform the users of updating the firmware. Most devices, which have been exposed to the Internet before



(a) Vulnerable Rate Trending of Router



(b) Vulnerable Rate Trending of IP camera



(c) Vulnerable Rate Trending of Printer

Fig. 1. Vulnerable Rate Trending of Three Specific Devices

the vulnerabilities were disclosed, still have high vulnerable rates. We notice that most devices' vulnerable rates begin to decrease since the newly produced devices with patched firmware have entered the market. In addition, the devices with automatic update mechanisms (e.g., Hikvision's IP cameras and HP's printers) can well defend against the N-days vulnerabilities. Overall, we conclude two root causes of differences in patching curves: 1) Whether the vendors have fixed the vulnerabilities in subsequent products in time; 2) Whether the vendors provide automatic update mechanisms to their products.

4.3 Geographical Difference

Since the IoT devices in our dataset are distributed all over the world, their vulnerability may have a connection with geography. Thus, we conduct an analysis to explore the relationship between the vulnerability results of IoT devices and their locations. We measure the number of vulnerable devices in each country, and for each kind of IoT device, we select the top-5 countries that contain the most vulnerable devices of this kind for further in-depth analysis. We obtain the geolocation of vulnerable IoT devices based on the information provided by the IoT search engine. The IoT search engine returns multiple kinds of information of IoT devices, including their geolocation according to several IP geolocation database providers. Table 7 shows the details of these selected countries, based on which we have the following observations.

First of all, we find the vulnerable routers are mainly concentrated to a limited number of countries. For the vulnerable routers of HUAWEI, most of them are located in Mexico. For the vulnerable routers of TP-Link, D-Link, ASUS, and MikroTik, it is surprising to observe that the United States are all included in the top-5 countries list of these four kinds of devices. When it comes to analyze the distribution of all the vulnerable routers, we find China has always been among the top-5 countries. We also observe that most of the vulnerable routers of D-Link, ASUS and MikroTik are all located in China.

Secondly, the ratio of the vulnerable IP cameras of the top-5 countries vary greatly from different vendors. For Hikvision, the vulnerable IP cameras are mainly located in China and Vietnam, accounting for more than 55% of all vulnerable devices. Then, the top-5 countries account for 84.9% of all vulnerable IP cameras from DAHUA. For Axis, the United States contains more than 60% of vulnerable IP cameras which is far beyond other countries. For Avtech, we find the top-5 countries are all located in Southeast Asia and Thailand has the most vulnerable IP cameras. Next, two European countries, France and Germany, account for 50% of all vulnerable IP cameras from Netwave.

Thirdly, the vulnerable printers from different vendors are mainly located in the United States. In addition to the United States, China has the second largest number of vulnerable printers, which is included to the top-5 list of HP, Canon, Brother, EPSON and Octoprint. We also find that the printers located in South Korea are in a very dangerous situation since a significant number of vulnerable printers of vendors (except Octoprint) are located in South Korea.

Then, we find most vulnerable mining devices are mainly located in China and the United States. For Claymore, China, the United States, Russia, Germany and India are the top-5 countries with the most vulnerable devices. For Antminer, we notice that 3 of 4 vulnerable devices are in China and the remaining one is in the United States.

Next, we find that only GE Healthcare has two vulnerable devices, both of which are located in Germany.

Finally, vulnerable ICS devices are only distributed in a few countries. All vulnerable devices from Rockwell Automation are located in three countries: the United States, Germany and France. Next, two vulnerable devices of WAGO are both in the United States. For Schneider Electric,

the United States has nearly 70% of vulnerable devices while there are 83 for Germany and 47 for China.

Overall, we have confirmed that the location has a significant relationship with the security of IoT devices. Several countries, e.g., the United States and China, contain much more vulnerable devices than others. Besides, the geographical distribution of vulnerable devices varies greatly between different kinds of IoT devices and different vendors.

5 Case Studies

In this section, we provide two case studies as important extensions to the study in Section 4. we revealed the in-depth findings based on two further experiments. First, we study how many IoT devices in our dataset are infected by the botnet. We then focus on the security of IoT devices' message center - MQTT servers.

5.1 Infected IoT Devices by Botnets

Our measurement results in Section 4 show that 385,060 (28.25%) IoT devices are still under the threat of at least one N-days vulnerability. Many of those vulnerable devices might have already been identified and infected by hackers before our research. Hence, we attempt to use NetworkScan Mon [43] and OpenData [44] provided by 360 NetLab [45] to identify vulnerable devices that have been previously infected.

OpenData provides us a set of the latest bot IPs that are identified by 360 NetLab, thus we first compare the IPs of vulnerable devices with these identified bot IPs. Through this comparison, we can preliminarily figure out part of the infected devices in our dataset. For the remaining devices, we use the NetworkScan Mon to identify whether they are infected or not. The NetworkScan Mon can detect active scanners on the Internet and capture more than 10k scanner IPs every day. Then, it adopts a neat way, which has not yet been disclosed to the public, to monitor their scan activities. The NetworkScan Mon provides us a dataset consisting of the device activities within 30 days. We query the remaining devices in the activity dataset provided by NetworkScan Mon, and label the device as infected if abnormal activities are observed (e.g., scanning the port 22 for more than 100 times). Combining the results returnd by OpenData and NetworkScan Mon, we finally identify the infected IoT devices among the vulnerable devices collected in Section 4.

Table 8 shows the statistics of the infected devices that are identified by us. We find that a large number of IoT devices are infected by botnets and all of them belong to router, IP camera, or printer. Printer has the least number of infected devices though it has the largest number of vulnerable devices. Router and IP camera are the primary targets of botnet because they could provide sufficient network conditions to DDoS attack. It is worthy noting that a part of infected devices do not suffer from N-days vulnerabilities, implying that the botnet may have other methods to attack the IoT devices. For the three specialized devices, though some of them are vulnerable to the N-days vulnerability attack, none of them has been infected by botnet. We speculate that the hackers have less interests in infecting specialized devices since the specialized devices are less exposed to the Internet and have weaker N-days vulnerability compared to the general-purpose IoT devices.

The botnets have the ability to hide themselves [11]: they can close the network services that are running on infected devices, e.g., Telnet and SSH. In addition, the botnets make the infected devices silent until they send control commands to these infected devices. Therefore, the analysis of our experiments is a conservative investigation of the real world security issues of IoT devices, which in practice might be severer.

5.2 Unsecured MQTT Servers

MQTT [7] is a machine to machine (M2M) connectivity protocol. An MQTT system consists of clients communicating with a server, where the server is also known as "MQTT server". Recently, millions of IoT devices, especially the smart homes, communicate frequently with MQTT servers. A large portion of devices in our dataset also have connection with the MQTT servers. However, the MQTT servers also suffer greatly from security threats.

We first identify 64,309 live MQTT servers exposed to the Internet, where most are deployed on popular Cloud Service Platforms, including Amazon Web Services (AWS), Google Cloud, and Microsoft Azure. However, due to the misconfiguration, a large number of these MQTT servers have no password protection, making the IoT devices connected to these unsecured MQTT servers easily controllable by hackers.

We analyze 14,477 unsecured MQTT servers that are deployed on five popular cloud service platforms, as shown in Table 9. AWS, Google Cloud and Microsoft Azure provide detailed IP ranges while Alibaba Cloud and Tencent Cloud do not. With the explicit IP ranges, we can collect the exposed MQTT servers that are deployed on AWS, Google Cloud, and Microsoft Azure. In addition, *Shodan* allows users to query the search engine by using the ISP information of the cloud services providers. By combining the two methods above, we collect a great number of exposed MQTT servers that are deployed on the five cloud services platforms.

We find that a great number of MQTT servers have no password protection. Therefore, we are able to connect to these unsecured MQTT servers directly and access sensitive data that are transferred between the IoT devices and MQTT servers. Alibaba Cloud has the most exposed MQTT servers and the highest vulnerable rate, which is defined as the proportion of unsecured MQTT servers out of all the exposed MQTT servers. For Alibaba Cloud, there are 6,936 MQTT servers exposed to the Internet and 92.6% of them have no password protection. AWS has 3,314 unsecured MQTT servers, which is sightly less than that of Alibaba Cloud. The other three cloud service platforms also have high vulnerable rates: more than 80% of the exposed MQTT servers have not set the password.

We notify all the affected cloud service providers and provide them with the IP address of the vulnerable servers. During the time we wrote this paper, AWS, Google Cloud, Tencent Cloud and Alibaba Cloud have responded to our report. Specifically, AWS, Google Cloud and Alibaba Cloud

TABLE 7
Top-5 Countries With the Most Vulnerable IoT Devices

	Vendors	Vulnerable Devices	Top-5 countries With the Most Vulnerable Devices				
Router	HUAWEI	885	Mexico (35.3%)	U.K. (23.3%)	China (16.6%)	Germany (9.4%)	Turkey (4.4%)
	TP-Link	2,193	U.S. (46.8%)	China (29.5%)	Turkey (4.4%)	Brazil (4.3%)	Russia (3.7%)
	D-Link	1,472	China (46.4%)	U.S.(21.1%)	Brazil (14.0%)	Russia (5.0%)	Italy (3.5%)
	ASUS	8,216	China (40.0%)	U.S. (25.6%)	Russia (11.9%)	Singapore (7.1%)	Sweden (6.3%)
	MikroTik	18,597	China (32.1%)	Russia (22.2%)	Brazil (11.1%)	Indonesia (10.2%)	U.S. (6.1%)
IP camera	Hikvision	18,104	China (37.1%)	Vietnam (21.4%)	U.S. (8.1%)	Mexico (7.1%)	Brazil (5.5%)
	DAHUA	13,499	Brazil (35.0%)	China (18.1%)	U.S. (12.9%)	Spain (10.5%)	Thailand (8.4%)
	Axis	35,137	U.S. (63.7%)	Germany (14.7%)	Japan (6.8%)	France (6.3%)	Mexico (2.9%)
	Avtech	46,571	Thailand (45.3%)	Indonesia (23.3%)	Vietnam (13.5%)	Malaysia (8.6%)	China (6.5%)
	Netwave	80,416	France (42.1%)	Germany (13.3%)	U.S. (11.4%)	China (10.7%)	Argentina (10.6%)
Printer	HP	25,518	U.S. (27.3%)	China (24.0%)	South Korea (18.3%)	Brazil (12.4%)	Canada (10.9%)
	Canon	3,398	U.S. (21.4%)	China (19.5%)	South Korea (18.5%)	Brazil (11.6%)	Chile (9.9%)
	Brother	65,938	U.S. (45.3%)	Canada (21.8%)	China (17.7%)	Germany (11.1%)	Japan (10.6%)
	EPSON	43,582	U.S. (27.2%)	South Korea (22.9%)	China (20.0%)	Japan (11.5%)	Germany (8.2%)
	SAMSUNG	73	U.S. (56.2%)	South Korea (43.8%)	-	-	-
	Octoprint	20,000	U.S. (35.6%)	Germany (35.3%)	U.K. (10.7%)	Italy (7.7%)	Spain (7.5%)
Mining Device	Claymore	1,034	China (54.4%)	U.S. (22.1%)	Russia (8.8%)	Germany (3.6%)	India (3.4%)
	Antminer	4	China (75.0%)	U.S. (25.0%)	-	-	-
Medical Device	GE Healthcare	2	Germany (100.0%)	-	-	-	-
ICS	Rockwell Automation	12	U.S. (50.0%)	Germany (25.0%)	France (25.0%)	-	-
	WAGO	2	U.S. (100.0%)	-	-	-	-
	Schneider Electric	407	U.S. (68.1%)	Germany (20.4%)	China (11.5%)	-	-

TABLE 8
Infected IoT Devices by botnets. #IDT represents the number of Infected Devices of the Total Devices. #IDV represents the number of Infected Devices of the Vulnerable Devices.

	Vendors	Total	Vulnerable Devices	#IDT / #IDV
Router	HUAWEI	100K	885	191 / 3
	TP-Link	100K	2,193	22 /19
	D-Link	100K	1,472	63 / 11
	ASUS	100K	8,216	215 / 167
	MikroTik	100K	18,597	409 / 394
IP camera	Hikvision	100K	18,104	526 / 513
	DAHUA	100K	13,499	108 / 72
	Axis	100K	35,137	636 / 580
	Avtech	100K	46,571	362 / 314
	Netwave	100K	80,416	85 / 73
Printer	HP	100K	25,518	341 / 201
	Canon	20K	3,398	2 / 2
	Brother	100K	65,938	94 / 91
	EPSON	50K	43,582	19 / 16
	Octoprint	20K	20,000	0 / 0
	SAMSUNG	20K	73	0 / 0
Mining Device	Claymore	30K	1,034	0 / 0
	Antminer	8K	4	0 / 0
Medical Device	GE Healthcare	300	2	0 / 0
	Softneta	92	0	0 / 0
	Dicoogle	14	0	0 / 0
ICS	Rockwell Automation	5K	12	0 / 0
	WAGO	500	2	0 / 0
	Schneider Electric	9K	407	0 / 0

TABLE 9
Unsecured MQTT Servers

Platforms	Total	Vulnerable Servers	
		First Test	Second Test
Amazon Web Services	4070	3314 (81.4%)	2942 (72.3%)
Alibaba Cloud	6936	6420 (92.6%)	6236 (89.9%)
Google Cloud	999	874 (87.5%)	801 (80.2%)
Microsoft Azure	1226	1030 (84.0%)	918 (74.9%)
Tencent Cloud	1246	1102 (88.4%)	1044 (83.8%)

acknowledge our work and send our security concern to the users of vulnerable servers while Tencent Cloud does not regard it as the security threat and ignores our report. We retest these unsecured MQTT servers after we have notified the cloud service platforms for two weeks. Though Microsoft Azure has not responded to our report yet, its vulnerable rate drops down by 9.1% from the first test. We positively infer that the Microsoft Azure has notified the users of vulnerable servers. The vulnerable rate of AWS has dropped down to 72.3% while the Alibaba Cloud, Google Cloud and Tencent Cloud still have more than 80% unsecured MQTT servers. All of these five cloud service platforms still have a great number of MQTT servers with no password protection even after we have notified them. In summary, the notification strategy of cloud service platforms and the limited security awareness of users need further attention for future development of IoT security.

6 Discussion

Ethics. In this work, we conduct N-days vulnerability test on 1,362,906 IoT devices, which may raise serious ethical concerns. In consideration of these potential hazards, we pay special attention to the legal and ethical boundaries. First, we collect all the data from IoT search engines, which is a legitimate manner and common practice to obtain data [11], [38], [46]. Second, we choose to leverage the firmware-fingerprinting method rather than the traditional PoC-Checking method to test the IoT devices. The firmware-fingerprinting method is an acceptable way to test online devices without ownership since it neither triggers nor exploits the vulnerability [40]. Finally, we are particularly careful to ensure the privacy of vulnerable devices. We report our results to the related vendors which help mitigate the security threats (as acknowledged by these vendors).

Limitations. There are two major limitations in our work. First, our vulnerable device detection method is based on a simple validation mechanism: whether the target device's firmware version is included in the affected firmware of the corresponding N-days vulnerabilities. Nevertheless, the correctness of the assumption highly depends on the correctness of the affected firmware versions claimed by the N-days vulnerabilities. Thus, there may exist the false-positive case in our final results due to the incorrectness claims of the N-days vulnerabilities. To solve this issue, we selected the N-days vulnerabilities whose affected firmware versions are confirmed by the vendors, based on which, we can eliminate most false-positive cases. Moreover, in consideration of the ethical issues and our large-scale dataset, it is the only acceptable way that can meet our evaluation goal. Though the PoC-Checking method has higher accuracy in identifying the vulnerable devices, we cannot perform the PoCs directly on the IoT devices without ownership.

Second, when we conduct the case study of detecting the IoT devices infected by botnets, it is hard to confirm whether the N-day vulnerability is the cause of infection since the default credential attack may also be a possible cause. In this case study, we try our best to conduct a more in-depth analysis about the security status of IoT devices. Our analysis does reveal the severe security issues of theses devices, call for stronger protection.

Future Work. Vulnerability notification is still an open research problem for vendors and researchers [11]. IoT devices lack a public communication channel such that we cannot reach the owners of vulnerable devices directly. Even if we provide detailed information about vulnerable devices to the related vendors, it is difficult for them to notify the vulnerable users. The traditional notification strategies, such as obtaining the information from WHOIS records, are not suitable for notifying IoT devices.

Moreover, there is still a lack of effective countermeasures to mitigate security threats to IoT devices. Yang et al. designed the multipath onion IoT gateways to hide the hackable smart home from remote attacks [47]. Sharma et al. presented a study of the Zero-day threats for IoT devices and proposed a context graph based framework for mitigating Zero-day attacks [24]. However, the applicability of two methods are limited since they cannot be generally extended to most IoT devices. Recently, deception defense [48], which is an emerging technology for cyber security, has been introduced to protect large-scale systems. Deception defense is designed to protect against the Zero-day and N-days vulnerability attack as well as default credential attack, which are also the prominent security challenges of IoT devices. Therefore, we plan to develop a system based on deception defense to protect the IoT devices in the future.

7 Related Work

IoT Search Engines Study. IoT search engines have become an important tool for researchers. Several studies have evaluated the ability of IoT search engines or leveraged them to analyze the security of IoT devices. Bodenheim et al. evaluated *Shodan*'s ability in searching PLC and proposed a potential method to defend against *Shodan* [49]. Genge et al. developed a *Shodan*-based vulnerability assessment tool which aims at assessing the automated and passive vulnerabilities of Internet-facing services [50]. Simon et al. conducted a contactless vulnerability analysis with the help of *Shodan* and *Google* [38]. Antonakakis et al. [11] and Hastings et al. [46] both use *Censys* to collect data for their experiments.

IoT Device Fingerprinting. Feng et al. [51] utilized the the banner of 17 industrial control protocols, e.g, Siemens S7 and BACnet, to identify cyber-physical system devices on the Internet. Li et al. [40] designed generating fine-grained fingerprints based on the subtle differences between the filesystems of various firmware images. They introduced the NLP to process the file contents to obtain the fingerprints.

Large-Scale Analysis of IoT Security. Several studies have been proposed to assess the security of IoT devices. Cui et al. developed an embedded device default credential scanner, and evaluated the default credential usage problem in embedded devices at a world-wide scale [52], [53]. Heninger et al. investigated the security of weak keys on a broad scale and discovered that insecure RNGs are in widespread use, leading to a significant number of vulnerable RSA and DSA keys [54]. Hastings et al. discovered the usage of a great number of weak keys [46]. Costin et al. evaluated the security of IoT devices at firmware level [22], [23]. Fernandes et al. found the vulnerabilities of emerging IoT frameworks [55], and proposed Flowfence [56] which offers data flow protection for these frameworks. In addition, the security of IoT applications has also become a popular research direction. Celik et al. presented SAINT, a static taint analysis tool for IoT applications, to track the sensitive information in IoT applications [26]. They further proposed SOTERIA to validate the safety, security, and functional properties of IoT applications [57]. Kumar et al. [25] conducted a large-scale analysis of IoT devices in real-world homes. They figured out the differences in IoT devices among regions. Their work mainly focuses on the IoT devices in the household that are not exposed to the Internet while we focus on the IoT devices exposed to the Internet. Alrawi et al. [58] evaluate the security of home-based IoT devices based on components analysis. They focus on the security of the IoT devices in five components while we focus on the N-days vulnerability problem in IoT devices. Granjal [59] analyzed existing protocols and mechanisms to secure communications in the IoT, as well as open research issues. Neshenko [60] focused on the ever-evolving IoT vulnerabilities and presented a first look on Internet-scale IoT exploitations. Humayed et al. [61] proposed an intensive literature review of CPS security based on their framework. They dedicated to providing a comprehensive overview of the state-of-the-art on CPS security.

8 Conclusion

In this paper, we present a large-scale empirical study on the vulnerability of $1,362,906$ deployed IoT devices lasting ten months. Furthermore, we perform the first systematic comparative study of five well-known IoT search engines and reveal the differences among them. We confirm that the N-days vulnerability attack is still a prominent security threat for IoT devices that need urgent attention. We also show a

broad view of vendors' defenses against these attacks and discover several limitations of these defenses.

Our work is a reminder to the public that the common security problems still seriously affect a significant number of IoT devices, and the existing countermeasures of vendors need further consideration and improvement.

ACKNOWLEDGMENT

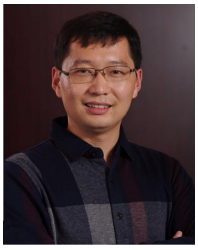
This work was partly supported by the National Key Research and Development Program of China under No. 2018YFB0804102, NSFC under No. 61772466, U1936215, and U1836202, the Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars under No. LR19F020003, the Zhejiang Provincial Key R&D Program under No. 2019C01055, and the Fundamental Research Funds for the Central Universities (Zhejiang University NGICS Platform). Wei-Han's work was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. Changting's work was partially supported by the State Key Laboratory of Information Security (Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093) (2020-MS-12), and the Zhejiang Provincial Natural Science Foundation No. LQ21F020010. Jingzheng's work was partially supported by NSFC under Grant 61772507. Pan's work was partially supported by NSFC under Grant 61972448. Liming's work was partially supported by NSFC under Grant 61872181.

REFERENCES

[1] "Shodan," https://www.shodan.io/, 2019.
[2] "Censys," https://www.censys.io/, 2019.
[3] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman, "A search engine backed by internet-wide scanning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2015, pp. 542–553.
[4] "Zoomeye," https://zoomeye.org/, 2019.
[5] "Fofa," https://fofa.so/, 2019.
[6] "Nti," https://nti.nsfocus.com/, 2019.
[7] "Mqtt," http://mqtt.org/, 2019.
[8] M. Hung, "Gartner insights on how to lead in a connected world," https://www.gartner.com/imagesrv/books/iot/iotEbook_digital.pdf, 2017.
[9] D. B. B. Herzberg and I. Zeifman, "Breaking down mirai: An iot ddos botnet analysis," https://www.incapsula.com/blog/malware-analysis-mirai-ddos-botnet.html, 2016.
[10] T. Sugawara, B. Cyr, S. Rampazzi, D. Genkin, and K. Fu, "Light commands: Laser-based audio injection on voice-controllable systems," 2019.
[11] 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, pp. 1092–1110.
[12] "Itu internet reports 2005: The internet of things," http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN021972.pdf, 2019.
[13] B. Beurdouche, K. Bhargavan, A. Delignat-Lavaud, C. Fournet, M. Kohlweiss, A. Pironti, P.-Y. Strub, and J. K. Zinzindohoue, "A messy state of the union: Taming the composite state machines of tls," in *2015 IEEE Symposium on Security and Privacy*. IEEE, 2015, pp. 535–552.
[14] Z. Durumeric, F. Li, J. Kasten, J. Amann, J. Beekman, M. Payer, N. Weaver, D. Adrian, V. Paxson, M. Bailey *et al.*, "The matter of heartbleed," in *Proceedings of the 2014 conference on internet measurement conference*. ACM, 2014, pp. 475–488.
[15] "Thingful," https://www.thingful.net, 2019.
[16] "Iot crawler," https://www.um.es/iotcrawler/, 2019.
[17] T. Heer, O. Garcia-Morchon, R. Hummen, S. L. Keoh, S. S. Kumar, and K. Wehrle, "Security challenges in the ip-based internet of things," *Wireless Personal Communications*, vol. 61, no. 3, pp. 527–542, 2011.
[18] C. Lyu, S. Ji, C. Zhang, Y. Li, W.-H. Lee, Y. Song, and R. Beyah, "MOPT: Optimized mutation scheduling for fuzzers," in *28th USENIX Security Symposium (USENIX Security 19)*, 2019, pp. 1949–1966.
[19] Y. Li, S. Ji, Y. Chen, S. Liang, W.-H. Lee, Y. Chen, C. Lyu, C. Wu, R. Beyah, P. Cheng *et al.*, "Unifuzz: A holistic and pragmatic metrics-driven platform for evaluating fuzzers," *arXiv preprint arXiv:2010.01785*, 2020.
[20] Y. Li, S. Ji, C. Lyu, Y. Chen, J. Chen, Q. Gu, C. Wu, and R. Beyah, "V-fuzz: Vulnerability prediction-assisted evolutionary fuzzing for binary programs," *IEEE Transactions on Cybernetics*, 2020.
[21] A. Cui, "The overlooked problem of 'n-day' vulnerabilities," https://www.darkreading.com/vulnerabilities---threats/the-overlooked-problem-of-n-day-vulnerabilities/a/d-id/1331348, 2018.
[22] 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.
[23] A. Costin, A. Zarras, and A. Francillon, "Automated dynamic firmware analysis at scale: a case study on embedded web interfaces," in *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*. ACM, 2016, pp. 437–448.
[24] V. Sharma, J. Kim, S. Kwon, I. You, K. Lee, and K. Yim, "A framework for mitigating zero-day attacks in iot," *arXiv preprint arXiv:1804.05549*, 2018.
[25] D. Kumar, K. Shen, B. Case, D. Garg, G. Alperovich, D. Kuznetsov, R. Gupta, and Z. Durumeric, "All things considered: An analysis of iot devices on home networks," in *28th USENIX Security Symposium (USENIX Security 19)*. Santa Clara, CA: USENIX Association, Aug. 2019, pp. 1169–1185. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/kumar-deepak
[26] Z. B. Celik, L. Babun, A. K. Sikder, H. Aksu, G. Tan, P. McDaniel, and A. S. Uluagac, "Sensitive information tracking in commodity iot," in *27th USENIX Security Symposium (USENIX Security 18)*, 2018, pp. 1687–1704.
[27] L. Lucero, "Unlocked doors," http://www8.hp.com/h20195/v2/GetPDF.aspx/4AA6-9485EEGB.pdf, 2018.
[28] B. BOTEZATU, "Ethereum os miners targeted by ssh-based hijacker," https://labs.bitdefender.com/2017/11/ethereum-os-miners-targeted-by-ssh-based-hijacker/, 2017.
[29] "Antminer," https://www.bitmain.com/, 2019.
[30] "Claymore," https://github.com/nanopool/Claymore-Dual-Miner, 2019.
[31] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 49–51, 2011.
[32] "Common vulnerabilities and exposures," http://cve.mitre.org, 2019.
[33] "Top 10 Free Open source PACS/DICOM Server Projects," https://medevel.com/10-open-source-pacs-dicom/.
[34] F. Valente, L. A. B. Silva, T. M. Godinho, and C. Costa, "Anatomy of an extensible open source pacs," *Journal of digital imaging*, vol. 29, no. 3, pp. 284–296, 2016.
[35] N. Goldenberg and A. Wool, "Accurate modeling of modbus/tcp for intrusion detection in scada systems," *International Journal of Critical Infrastructure Protection*, vol. 6, no. 2, pp. 63–75, 2013.
[36] G. F. Lyon, *Nmap network scanning: The official Nmap project guide to network discovery and security scanning*. Insecure, 2009.
[37] Z. Durumeric, E. Wustrow, and J. A. Halderman, "Zmap: Fast internet-wide scanning and its security applications." in *USENIX Security Symposium*, vol. 8, 2013, pp. 47–53.

[38] K. Simon, C. Moucha, and J. Keller, "Contactless vulnerability analysis using google and shodan," *Journal of Universal Computer Science*, vol. 23, no. 4, pp. 404–430, 2017.
[39] A. Cui, M. Costello, and S. Stolfo, "When firmware modifications attack: A case study of embedded exploitation," 2013.
[40] Q. Li, X. Feng, R. Wang, Z. Li, and L. Sun, "Towards fine-grained fingerprinting of firmware in online embedded devices," in *INFOCOM 2018*, 2018.
[41] "Exploit database," https://www.exploit-db.com/, 2019.
[42] "Seebug," https://www.seebug.org/, 2019.
[43] "Networkscan mon," https://scan.netlab.360.com/, 2019.
[44] "Opendata," https://data.netlab.360.com/, 2019.
[45] "360 netlab," http://netlab.360.com/, 2019.
[46] M. Hastings, J. Fried, and N. Heninger, "Weak keys remain widespread in network devices," in *Proceedings of the 2016 Internet Measurement Conference*. ACM, 2016, pp. 49–63.
[47] L. Yang, C. Seasholtz, B. Luo, and F. Li, "Hide your hackable smart home from remote attacks: The multipath onion iot gateways," in *European Symposium on Research in Computer Security*. Springer, 2018, pp. 575–594.
[48] D. Fraunholz, S. D. Anton, C. Lipps, D. Reti, D. Krohmer, F. Pohl, M. Tammen, and H. D. Schotten, "Demystifying deception technology: A survey," *arXiv preprint arXiv:1804.06196*, 2018.
[49] R. Bodenheim, J. Butts, S. Dunlap, and B. Mullins, "Evaluation of the ability of the shodan search engine to identify internet-facing industrial control devices," *International Journal of Critical Infrastructure Protection*, vol. 7, no. 2, pp. 114–123, 2014.
[50] B. Genge and C. Enăchescu, "Shovat: Shodan-based vulnerability assessment tool for internet-facing services," *Security and communication networks*, vol. 9, no. 15, pp. 2696–2714, 2016.
[51] X. Feng, Q. Li, H. Wang, and L. Sun, "Characterizing industrial control system devices on the internet," in *2016 IEEE 24th International Conference on Network Protocols (ICNP)*. IEEE, 2016, pp. 1–10.
[52] A. Cui, Y. Song, P. V. Prabhu, and S. J. Stolfo, "Brave new world: Pervasive insecurity of embedded network devices," in *International Workshop on Recent Advances in Intrusion Detection*. Springer, 2009, pp. 378–380.
[53] A. Cui and S. J. Stolfo, "A quantitative analysis of the insecurity of embedded network devices: results of a wide-area scan," in *Proceedings of the 26th Annual Computer Security Applications Conference*. ACM, 2010, pp. 97–106.
[54] N. Heninger, Z. Durumeric, E. Wustrow, and J. A. Halderman, "Mining your ps and qs: Detection of widespread weak keys in network devices." in *USENIX Security Symposium*, vol. 8, 2012, p. 1.
[55] E. Fernandes, J. Jung, and A. Prakash, "Security analysis of emerging smart home applications," in *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2016, pp. 636–654.
[56] E. Fernandes, J. Paupore, A. Rahmati, D. Simionato, M. Conti, and A. Prakash, "Flowfence: Practical data protection for emerging iot application frameworks," in *25th USENIX Security Symposium (USENIX Security 16)*, 2016, pp. 531–548.
[57] Z. B. Celik, P. McDaniel, and G. Tan, "Soteria: automated iot safety and security analysis," in *2018 USENIX Annual Technical Conference (USENIX ATC 18)*, 2018, pp. 147–158.
[58] O. Alrawi, C. Lever, M. Antonakakis, and F. Monrose, "Sok: Security evaluation of home-based iot deployments," in *2019 IEEE Symposium on Security and Privacy (SP)*, 2019, pp. 1362–1380.
[59] J. Granjal, E. Monteiro, and J. Sá Silva, "Security for the internet of things: A survey of existing protocols and open research issues," *IEEE Communications Surveys Tutorials*, vol. 17, no. 3, pp. 1294–1312, 2015.
[60] N. Neshenko, E. Bou-Harb, J. Crichigno, G. Kaddoum, and N. Ghani, "Demystifying iot security: An exhaustive survey on iot vulnerabilities and a first empirical look on internet-scale iot exploitations," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2702–2733, 2019.
[61] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-physical systems security—a survey," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1802–1831, 2017.

Binbin Zhao is currently a Ph.D. student in the School of Electrical and Computer Engineering at Georgia Tech. He received the BS degree from the School of Computer Science at Zhejiang University in 2018. His research interest includes IoT security, CAPTCHA, and Adversarial Learning.



Shouling Ji is a ZJU 100-Young Professor in the College of Computer Science and Technology at Zhejiang University and a Research Faculty in the School of Electrical and Computer Engineering at Georgia Institute of Technology. He received a Ph.D. in Electrical and Computer Engineering from Georgia Institute of Technology, a Ph.D. in Computer Science from Georgia State University. His current research interests include AI Security, Data-driven Security, Privacy and Data Analytics. He is a member of IEEE and ACM and was the Membership Chair of the IEEE Student Branch at Georgia State (2012- 2013).

Wei-Han Lee is a research staff member in IBM T.J. Watson Research Center. His current research interest includes AI security, big data privacy, and distributed AI. He studied at Princeton University for Ph.D. in Electrical Engineering, and at National Taiwan University for Bachelor in Physics and Electrical Engineering.

Changting Lin received the Ph.D's degree in computer science from the Zhejiang University in 2018. His is currently a post-doctoral of Big data statistics method and application at Zhejiang Gongshang University, China. His research interests include IoT, network security and Blockchain.

Haiqin Weng is with the Ant Group, Hangzhou, China. Her research interests include AI security, anomaly detection, and machine learning. She received her Ph.D. degree in College of Computer Science and Technology at Zhejiang University in 2019, and received her BS degree from South China University of Technology in 2014.

Jingzheng Wu received his Ph.D. degree in computer software and theory from the Institute of Software, Chinese Academy of Sciences, Beijing, in 2012. He is a research professor at the Institute of Software, Chinese Academy of Sciences, Beijing. His primary research interests include system security, vulnerability detection, covert channels.

Pan Zhou (S'07–M'14-SM'20) is currently an associate professor and Ph.D. advisor with Hubei Engineering Research Center on Big Data Security, School of Cyber Science and Engineering, Huazhong University of Science and Technology (HUST), Wuhan, P.R. China. He received his Ph.D. in the School of Electrical and Computer Engineering at the Georgia Institute of Technology (Georgia Tech) in 2011, Atlanta, USA. He received his B.S. degree in the Advanced Class of HUST, and a M.S. degree in the Department of Electronics and Information Engineering from HUST, Wuhan, China, in 2006 and 2008, respectively. He held honorary degree in his bachelor and merit research award of HUST in his master study. He was a senior technical member at Oracle Inc., America, during 2011 to 2013, and worked on Hadoop and distributed storage system for big data analytics at Oracle Cloud Platform. He received the "Rising Star in Science and Technology of HUST" in 2017. He is currently an associate editor of IEEE Transactions on Network Science and Engineering. His current research interest includes: security and privacy, big data analytics and machine learning, and information networks.

Liming Fang received the Ph.D. degree in Computer Science from Nanjing University of Aeronautics and Astronautics in 2012, and has worked as a postdoctoral fellow in the information security from City University of Hong Kong. Now, he is an associate professor at the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics. His current research interests include cryptography and information security. Liming Fang has published more than 50 paper in his field, including IEEE TDSC, IEEE TIFS, Theoretical Computer Science, Designs Codes and Cryptography, Information Sciences, etc.

Raheem Beyah , a native of Atlanta, GA, serves as Georgia Tech's Vice President for Interdisciplinary Research, Executive Director of the Online Masters of Cybersecurity program (OMS Cybersecurity), and is the Motorola Foundation Professor in School of Electrical and Computer Engineering. He is also Co-Founder of Fortiphyd Logic , Inc. Raheem received his Bachelor of Science in Electrical Engineering from North Carolina A&T State University in 1998. He received his Masters and Ph.D. in Electrical and Computer Engineering from Georgia Tech in 1999 and 2003, respectively. His research interests include Network security and monitoring, Cyber-physical Systems Security, Network traffic characterization and performance, and Critical infrastructure security. He received the National Science Foundation CAREER award in 2009 and was selected for DARPA's Computer Science Study Panel in 2010. He is a member of AAAS, ASEE, a lifetime member of NSBE, a senior member of IEEE, and an ACM Distinguished Scientist.