Regular Paper

Disposable Botnets: Long-term Analysis of IoT Botnet Infrastructure

Rui Tanabe[1,a)] Tsuyufumi Watanabe[1,3,b)] Akira Fujita[2,c)] Ryoichi Isawa[2,d)]
Carlos Gañán[4,e)] Michel van Eeten[4,f)] Katsunari Yoshioka[1,g)] Tsutomu Matsumoto[1,h)]

Received: December 7, 2021, Accepted: June 14, 2022

Abstract: Large botnets made up of Internet-of-Things (IoT) devices have a steady presence in the threat landscape since 2016. However, it has not explained how attackers maintain control over their botnets. In this paper, we present a long-term analysis of the infrastructure of IoT botnets based on 36 months of data gathered via honeypots and the monitoring of botnet infrastructure. We collected 64,260 IoT malware samples, 35,494 download servers, and 4,736 C&C servers during 2016 to 2021. Not only are most binaries distributed for less than three days, but the connection of bots to the rest of the botnet is also short-lived. To reach the C&C server, the binaries typically contain only a single hard-coded IP address or domain. Long-term dynamic analysis finds no mechanism for the attackers to migrate the bots to a new C&C server. Although malware binaries that use domain names to connect to their C&C servers increased in 2020, the C&C servers themselves have a short lifespan and this tendency has not changed. The picture that emerges is that of highly disposable botnets. IoT botnets are reconstituted from scratch all the time rather than maintained.

Keywords: Internet-of-Things, IoT malware binary, C&C server, IoT honeypot

1. Introduction

The rise of the Internet-of-Things (IoT) is causing dramatic changes in the Internet ecosystem. Billions of heterogeneous devices are being connected to the Internet. Smart meters have been rolled out, traffic management is enhanced with road sensors and smart traffic lights, entire manufacturing plants are monitored over the Internet, automated homes can be controlled remotely and are equipped with connected devices like fridges, washing machines, and security cameras—the list goes on and on.

The overall security of IoT devices has not kept up with these developments. Especially in the consumer space, devices are being scanned and hacked at scale [1]. Morteza et al. [2] recently found more than 400,000 exploited IoT devices across 350 IoT botnets. Different IoT malware families, such as Bashlite [3], Mirai [4], and Tsunami (a.k.a., Kaiten) [5] have enabled attackers to launch distributed denial-of-service (DDoS) attacks [6], cryptojacking campaigns [7] and other forms of cybercrime [8], [9].

Prior work has uncovered three key aspects of IoT botnets. First, malware binaries are seen only briefly, often for less than one day, and that the download servers of the binaries are also short lived [10]. Second, it has been shown that command and control (C&C) servers stay online only a bit longer, typically for a handful of days, at least for Bashlite and Mirai [11]. Besides, malware rely on hard-coded IP addresses for C&C callback [12]. Third, download servers and C&C servers are concentrated in few ASes associated with cloud providers, at least for Bashlite and Mirai botnets [13]. Similarly, the P2P IoT botnet Hajime [14], [15] is also short lived and has a particularly heavy concentration in a small number of countries [16]. While these works provide fascinating glimpses, we still lack a coherent picture of IoT botnet infrastructure. Prior work leaves unanswered how attackers maintain their botnet in light of all these volatile components.

This paper presents a longer-term study of IoT botnet infrastructure. We investigate how malware binaries connect to their C&C servers and how attackers update their binaries or how they refresh C&C server information. We show how binaries, download servers and C&C servers are related to each other and evolve over time. Our analysis focuses on Bashlite, Mirai and Tsunami, because they are major samples on our IoT Honeypot It is based on a much larger dataset than prior work, containing 64,260 IoT malware samples, 35,494 download servers, and 4,736 C&C servers. These were captured over a period of 36 months of IoT honeypot operations (October 2016–December 2017, October 2018–May 2019, and January 2020–January 2021). In short, we find a range of botnets that take on a highly disposable form, something never seen before in Windows-based botnets. We

1 Yokohama National University, Yokohama, Kanagawa 240–8501, Japan
2 National Institute of Information and Communications Technology, Koganei, Tokyo 184–8795, Japan
3 FUJISOFT Incorporated, Yokohama, Kanagawa 231–8008, Japan
4 Delft University of Technology, Delft, Netherlands
a) tanabe-rui-xj@ynu.ac.jp
b) watanabe-tsuyufumi-vj@ynu.jp
c) a.fujita@nict.go.jp
d) isawa@nict.go.jp
e) c.hernandezganan@tudelft.nl
f) m.j.g.vaneeten@tudelft.nl
g) yoshioka@ynu.ac.jp
h) tsutomu@ynu.ac.jp

The primary version of this work was presented at ARES 2020 [17].

show IoT botnets are reconstituted from scratch all the time rather than maintained. The concentration in certain cloud providers might be the effect of abusing IP address allocation practices of these providers in order to get fresh IPs for the constantly changing C&C servers. Although the overall tendency has not changed, we note that some attackers attempt to hide their C&C connection in 2020. These disposable botnets creates new challenges and opportunities in the combat against IoT botnets. Our main contributions are as follows:

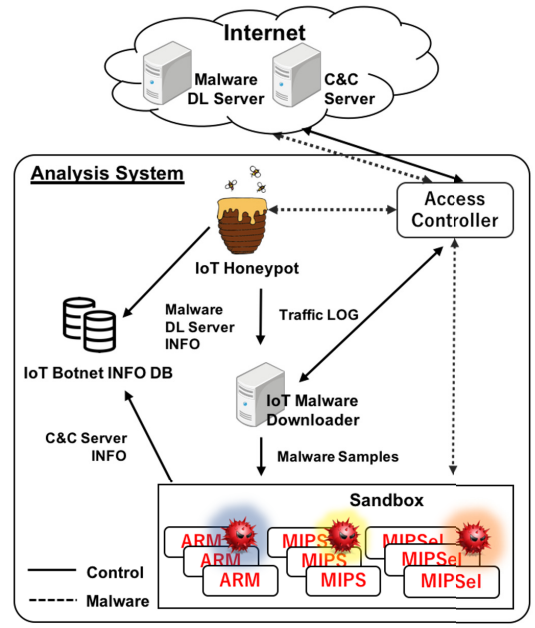
- About 85% of all binaries are distributed for less than 5 days. The binaries have only a fleeting connection to their infrastructure, as they contain only one or few C&C IP addresses. C&C servers, in turn, are also online for just a few days. Long-term dynamic analyses found no attempts by the C&C server to update the binary of the bot before the connection is lost. All of this means that the attackers treat the bots, and in fact the whole botnet, as "disposable".
- Download servers and C&C servers are mostly located in cloud providers, similar to traditional Windows-based botnets. This indicates attackers are not using compromised IoT devices to control their botnets. Around 70% of the download servers and C&C servers are active for 5 days or less. This pattern has not changed over the past years.
- There are a few Autonomous Systems (ASes) where a large portion of the botnet infrastructure is concentrated. A single AS of a cloud provider accounts for 39% of the total number of detected C&C servers. This concentration is present independently of the malware family. C&C servers in such ASes are not used at the same time and their IP addresses are seen only in subsequent periods. We speculate that the attackers might be abusing the provisioning of IP addresses to virtual servers, thereby easily sourcing short-lived addresses that can circumvent blacklisting.
- In 2020, C&C servers themselves were still short-lived and that the tendency has not changed for years. However, malware binaries that use domain names for their C&C connection slightly increased. Domain names that had strings indicating C&C (for example, `cnc` and `c2`) in their names decreased. A few binaries were using Tor nodes for further connections. All of these evidence indicates that some attackers are trying to hide their C&C connection.

The remainder of this paper is structured as follows. In Section 2, we outline the structure of our monitoring and analysis system. In Section 3, we present descriptive results of the data captured by the monitoring system. In Section 4 we analyze botnet infrastructures and describe how attackers maintain their infrastructure.

In Section 5, we discuss the limitations of our work. In Section 6, we explain the related work. Finally, we conclude in Section 7.

2. Methodology

As malicious software which targets IoT devices continues to grow, it has allowed attackers to create IoT botnets. These IoT malware typically infect devices that provide SSH or Telnet network services, by exploiting default passwords, weak credentials,



SPECIFICATION				
	OS	CPU	Memory	HD
Analysis System	Ubuntu 14.04	Intel®CPU E3-1240 V2@3.40GHz*4	16G	2T
ARM/MIPS/MIPSEL	Debian 7.3	QEMU	124MB	24GB

Fig. 1 Structure of analysis system.

or simply brute-forcing techniques. Once compromised, malicious binaries are downloaded to the device for enrollment into the IoT botnet. The attacker then gains control of the device and can send commands through command and control (C&C) servers or peer-to-peer (P2P) networks. To understand the infrastructure used by different IoT botnets, we started by observing IoT-related attacks as captured by a hybrid honeypot scheme that combines low and high interaction.

2.1 Monitoring Infrastructure

Inspired by the work of IoTPOT [18], we designed a similar honeypot. Instead of virtual environments, we used bare-metal IoT devices as high-interaction honeypots and performed a study in a greater detail. We also used a sandbox environment for dynamically analyzing malware collected by the honeypot; and for some cases, we conducted a static analysis. **Figure 1** illustrates the overview structure of our analysis system. The system consists of five components, as follows:

IoT Honeypot: Compared to IoTPOT [18], we have extended our honeypot with additional services such as known abused HTTP front-ends, the CPE WAN Management Protocol (CWMP) [19], a backdoor of the Netis router, and the remote access setup service of several IP cameras. In case of high-interaction honeypots, we used four different bare-metal IoT devices (two WiFi storage devices, one router, and one IP camera) for observation. Once an attacker logs into the honeypot and obtains a privileged system shell, we recorded the system interactions, including shell commands.

IoT Malware Downloader: The IoT malware downloader extracts malware download commands from shell command sequence which are observed by IoT honeypot. We download IoT

malware samples as soon as the download command is observed.

Sandbox: The sandbox is composed of virtual machines and bare-metal devices. The sandbox uses QEMU for emulating devices with the three most prevalent CPU architectures of the collected samples: ARM, MIPS, and MIPSEL, on which a Linux OS (Debian) is running. The bare-metal sandbox uses four types of physical devices, which are the same as the IoT Honeypot (ARM, MIPS, and MIPSEL). These devices are chosen as they have indeed been identified as infected devices by our honeypot and are typical off-the-shelf, low-cost hardware. We note that bare-metal sandboxes run the risk of malfunctioning, as they run untrusted malware binaries. However, Authors in Ref. [20] reported that existing IoT malware is non-persistent and thus can be easily removed by rebooting the infected device. Therefore, we reboot the devices after each malware execution and compare files and processes in the devices to see if any changes remain after the reboot.

Access Controller: The access controller controls the traffic between the Internet and the IoT Honeypot, Malware Downloader, and Sandbox. It also forwards inbound traffic such as telnet, to honeypot for passive monitoring. On the other hand, it filters out dangerous outbound attacks.

IoT Botnet INFO DB: Every information such as hash codes of IoT malware binaries, DL servers, C&C servers and the relationship among them are stored into a database.

2.2 Measurement Period

Our monitoring infrastructure evolved over time to capture the evolution of malware families targeting different IoT devices. Our aim was to capture new IoT botnets targeting different devices.

To account for these changes in the honeypot, we report the analyses of 2 non-overlapping measurements periods. The first measurement period comprises from October 2016 to December 2017, while the second period goes from October 2018 to May 2019. During the period from December 2017 to October 2018, the honeypot infrastructure was upgraded and moved from a research network to a commercial ISP network to increase the deception capabilities of the honeypot. After the second period, we continued our measurement and used a part of the data for dynamic analysis.

2.3 Dynamic Analysis

During both measurement periods, we analyzed dynamically all the binaries collected by the hybrid honeypot. We conducted 2 different types of dynamic analysis, i.e., short-term and long-term analyses. The aim of the short-term analysis is to extract C&C server IP addresses and domain names, while the long-term analysis aimed at understanding how the bot master controlled/updated their bots. Hence the short-term analysis only last a couple of minutes until we detected the first communication attempt from the sandbox to the C&C server. On the other hand, the long-term analyses were run for several days (up to 7 days).

As the functionality of the binaries is independent of the architectures, for pragmatic issues, we selected a MIPS bare-metal device for the short-term analysis (most stable sandbox) and a MIPSEL sandbox for the long-term analyses. In the case of the short-term analysis, we analyzed all malware binaries right after they were captured to increase the chances of the C&C server still being up and running. The short-term analysis was conducted with isolated sandboxes without any Internet connection.

The long-term analysis was conducted the period from October 2019 to November 2019. During the period, we examined the results of each short-term analysis right after the analysis concluded to see if the binary tried to connect to C&C server. If so, we sent the binary to the long-term analysis immediately. We performed our analysis in parallel and tried not to occupy the sandbox analysis system. For ethical reasons, we filter out scan packets, DDoS attacks, and other attack traffic, while still allowing network connections to the C&C servers. We closely monitored the traffic until 24 hours after the connection to C&C server was lost, to see if there was any attempt to re-establish a connection to the C&C infrastructure or any backup servers. Because of the manual effort involved, we could only analyze 50 malware samples during the second measurement period. We have to leave the analysis with more exhaustive samples for future work.

3. Discovering IoT Botnets

As a first step to understand IoT botnet infrastructures, we analyze the distribution of malware binaries and the global distribution of download servers and C&C servers.

3.1 Malware Binary

We deployed our IoT honeypot at 808 IP addresses distributed across three countries (henceforth measurement one). As shown in the left side of **Table 1**, one sensor was located in the Netherlands within a /24 network, 5 sensors were deployed in Taiwan within /26 network and 2 additional sensors in Japan in different /24 networks. The data was collected from October 2nd 2016 to December 2nd 2017. During the period, we collected 50,026 unique malware binaries. To understand the type of devices that were targeted, we analyzed the CPU architecture of each IoT malware. The collected IoT malware had a great diversity in CPU architecture but the main parts were ARM (22.87%), Intel 80386 (15.72%), MIPS (12.38%), and MIPSel (11.07%). The reason may be that the attackers attempted to infect as many IoT devices as possible. In many cases, malware binaries for multiple CPU architectures were downloaded and executed. We then identified malware families of Bashlite, Mirai, and Tsunami. Hereafter,

Table 1 IoT honeypot sensor.

Measurement one			Measurement two			Measurement three		
Location	**#IP Addresses**	**Period**	**Location**	**#IP Addresses**	**Period**	**Location**	**#IP Addresses**	**Period**
Netherlands	253	Dec, 2016–Dec, 2017	Japan	130	Oct, 2018–May, 2019	Japan	130	Jan, 2020–Jan, 2021
Taiwan	5 * 63	Dec, 2016–May, 2017		-		-		
Japan	140	Oct, 2016–Dec, 2017		-		-		
Japan	130	Nov, 2016–Dec, 2017		-		-		

we used Anti-virus software Dr.WEB's [21] to label each binary. While the AV labels provide some information, the distinction between Bashlite, Mirai, and Tsunami is not clearcut due to code reuse across families. We do indeed see that much of the behavior across these families are very similar and we describe most of our findings as pertaining to all three families.

The left side of **Table 2** summarizes the number of malware binaries that belonged to each malware family, with the associated number of download (DL) servers (measurement one). We note that Hajime is a P2P based IoT malware and the infected host becomes download server and C&C server.

To further understand botnet infrastructures, we also deployed our IoT honeypot during a relatively recent time period (measurement two). As shown in the right side of Table 1, all sensors were located in Japan. The data was collected from October 2018 to May 2019. During the period, we collected 9,858 unique malware binaries. Similarly to the previous measurement, the collected IoT malware had a great diversity in CPU architecture. However, during this period of observation, we did not use any ARM device for the high-interaction honeypot. The location of the sensor and the number of IP addresses are different. Therefore, the portion of malware binaries are slightly different. Out of these binaries, we identified malware families of Bashlite, Mirai, and Tsunami. The right side of Table 2 summarizes the number of malware binaries with the number of download (DL) servers (measurement two).

3.2 Malware Download Server

After capturing malware binaries, we analyzed the global distribution of malware download servers. For measurement one, we identified 23,341 unique IP addresses, distributed across 145 countries and 1,405 ASes. The left side of **Table 3** shows the results for measurement one: the top 10 ASes with the highest number of download servers for each malware family. We used GeoIP2Country [22] to identify AS number and country code (CC). We used IP2Location [23] "Usage Type" to identify the AS usage. ASes are colored according to their types (Data Center/Web hosting/Transit, Broadband ISP, Mobile ISP, and Governments). Although there is some variance in the ranking, the main locations of download servers of Bashlite, Mirai and Tsunami were US and European countries. Most download servers were using "Data Center/Web Hosting/Transit" services. We also looked into malware binaries downloaded from the same DL server with a case study of DL server "XXX.239.72.250". In **Fig. 2**, gray points are malware binaries downloaded from the DL server. Pink squares are download dates. Binaries linking to the square means they are downloaded in that date. In this case, among 1,202 binaries downloaded in 2017-09-09, only 199 of them were downloaded in 2017-09-10. The figure shows that malware binaries are indeed frequently updated within a single download server.

Table 2 Summary of malware binaries & download servers.

	Measurement one		Measurement two	
Family	**#Binaries**	**#DL**	**#Binaries**	**#DL**
Bashlite	43,855	1,358	332	175
Mirai	4,275	440	9,171	3,042
Tsunami	602	52	63	29
Other	1,294	21,491 *	292	9,132 *
Total	50,026	23,341 *	9,858	12,177 *

* These numbers include 21,134 Hajime infected hosts in Measurement one and 9,004 Hajime infected hosts in Measurement two, respectively.

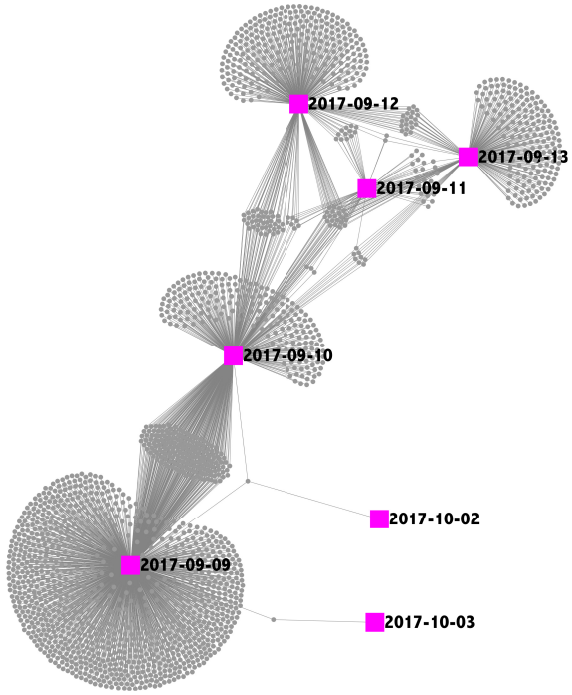


Fig. 2 Relationship of malware binaries and download dates.

Table 3 Top 10 ASes of download servers per malware family.

	Measurement one									Measurement two								
	Bashlite			Mirai			Tsunami			Bashlite			Mirai			Tsunami		
	AS	**CC**	**#DL**	**AS**	**CC**	**#DL**	**AS**	**CC**	**#DL**	**AS**	**CC**	**#DL**	**AS**	**CC**	**#DL**	**AS**	**CC**	**#DL**
1	23033	US	167	12876	FR	50	20473	US	7	14061	US	43	14061	US	933	14061	NL	5
2	20473	US	117	31034	IT	31	12876	FR	5	60144	NL	12	14061	NL	192	14061	US	4
3	31034	IT	108	20473	US	29	60781	NL	4	54290	US	11	51659	RU	94	53667	US	4
4	36352	US	97	43350	NL	20	31034	IT	4	31034	IT	11	60144	NL	78	51659	RU	2
5	33387	US	53	36352	US	16	23033	US	4	53667	US	10	20473	US	74	12876	FR	2
6	53755	US	52	29073	SC	10	44812	RU	3	14061	NL	9	54290	US	71	53667	LU	1
7	393406	US	46	393406	US	9	43350	NL	3	51659	RU	8	14061	SG	71	14061	SG	1
8	200039	GB	40	49981	NL	8	36352	US	2	3842	US	5	31034	IT	70	51167	DE	1
9	43350	NL	38	4766	KR	7	33387	US	2	24806	CZ	4	53667	US	59	200651	SC	1
10	49349	NL	27	47381	HU	7	62282	LT	1	43350	NL	3	14061	GB	44	204725	UA	1

Data Center/Hosting/Transit | Broadband ISP

Table 4 Top 10 ASes of C&C Servers per malware family (CPU Arch. MIPS).

	Measurement one									Measurement two								
	Bashlite			Mirai			Tsunami			Bashlite			Mirai			Tsunami		
	AS	CC	#C&C	AS	CC	#C&C	AS	CC	#C&C	AS	CC	#C&C	AS	CC	#C&C	AS	CC	#C&C
1	23033	US	128	31034	IT	7	31034	IT	5	14061	US	38	14061	US	593	12876	FR	2
2	31034	IT	100	12876	FR	6	20473	US	4	60144	NL	15	14061	NL	111	14061	US	1
3	20473	US	84	49981	NL	2	43350	NL	1	14061	NL	6	60144	NL	71	14061	NL	1
4	36352	US	68	44812	RU	2	24961	FR	1	53667	US	6	51659	RU	59	31034	IT	1
5	393406	US	38	43350	NL	2	14061	NL	1	54290	US	5	54290	US	54	53667	US	1
6	43350	NL	35	29073	SC	2	-	-	-	31034	IT	4	31034	IT	46	200185	IT	1
7	53755	US	30	200019	MD	2	-	-	-	3842	US	3	20473	US	44	51659	RU	1
8	200039	GB	30	197226	PL	2	-	-	-	14061	SG	3	14061	SG	38	51731	CZ	1
9	33387	US	25	9605	JP	1	-	-	-	200185	IT	3	53667	US	37	197695	RU	1
10	60781	NL	21	8896	NO	1	-	-	-	51659	RU	3	14061	DE	36	-	-	-

Data Center/Hosting/Transit | Mobile ISP | Government

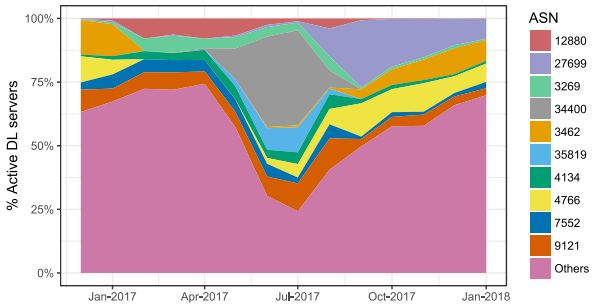


Fig. 3 AS transition of top 10 Hajime download servers.

We then analyzed the global distribution of malware download servers for measurement two. We identified 12,177 unique IP addresses, distributed across 131 countries and 1,957 ASes. The right side of Table 3 shows the top 10 ASes of in terms of download servers. Again we used IP2Location to identify AS usage and the coloring rules are the same. Likewise, the main locations of download servers were US and European countries and most of the download servers were using "Data Center/Web Hosting/Transit" services. Compared to the first measurement, many servers were located in AS14061. This address range belongs to a cloud service provider (Digital Ocean) and we further analyze these servers in Section 4.6. This finding confirmed prior work that also found that malware download servers were concentrated in cloud providers [10], [13].

Hajime downloads its binaries and transmits commands through its own P2P network of infected devices [16]. In case of Hajime, which is a P2P botnet, malware binaries were downloaded not from particular servers but from other members of the botnet. Thus, the sources of malware download were mostly end users of ISP service. We illustrate the monthly AS types transition of Top 10 Hajime download servers in **Fig. 3**. Interestingly, the ratio of Top 10 countries were dramatically varied. For example, download servers in AS34400 started from May 2017 and disappeared August 2017. This result indicates that Hajime's targets were continuously changing.

3.3 Command and Control Server

To extract C&C server information, we performed dynamic analysis in a sandbox environment. We focused on MIPS binaries due to the prevalence of binaries targeting this CPU architecture. We executed them in a sandbox environment for five minutes right after a new malware sample was captured by the IoT honeypot. We observed its behavior and identified their C&C server's IP addresses. Most of the samples connected to their C&C server without domain resolution. However, we discovered that responses from their C&C servers only last for a short period. After their connection has ended, these binaries kept connecting to their C&C server. None of the samples tried to connect to another C&C server. This indicates that malware binaries do not have such robustness in C&C connection.

During measurement one, we identified 650 unique IP addresses, distributed across 36 countries and 106 ASes. We then identified C&C servers corresponding to three different malware families: 543 Bashlite, 41 Mirai, and 12 Tsunami. The left side of **Table 4** shows the AS information of Top 10 C&C servers. The overall trend is the same as with download servers, most of them were located in US and European countries, the most AS types were "Data Center/Web Hosting/Transit".

In case of measurement two, we identified 2,302 unique IP addresses, distributed across 41 countries and 165 ASes. We then observed C&C servers corresponding to the three malware families: 137 Bashlite, 2,176 Mirai, and 10 Tsunami. The right side of Table 4 shows the AS information of Top 10 C&C servers. It is the same with download servers: most C&C servers were located in US and European countries, most ASes were "Data Center/Web Hosting/Transit", and most servers were located in AS14061. As in prior work, C&C servers are concentrated in cloud providers [11], [13].

4. Disposable Botnet Infrastructure

To further understand the botnet infrastructure, we estimate the update frequency of malware binaries, with their download servers and C&C servers. We then analyze the connectivity to their C&C servers. Finally, we investigate their relationships and summarize how attackers maintain their infrastructure.

4.1 Binary Update Frequency

Among malware binaries captured during measurement one and two, in this section, we focus on binaries that were targeting MIPS CPU architecture. We shed light into three famous malware families: Bashlite, Mirai, and Tsunami (henceforth dataset). **Ta-**

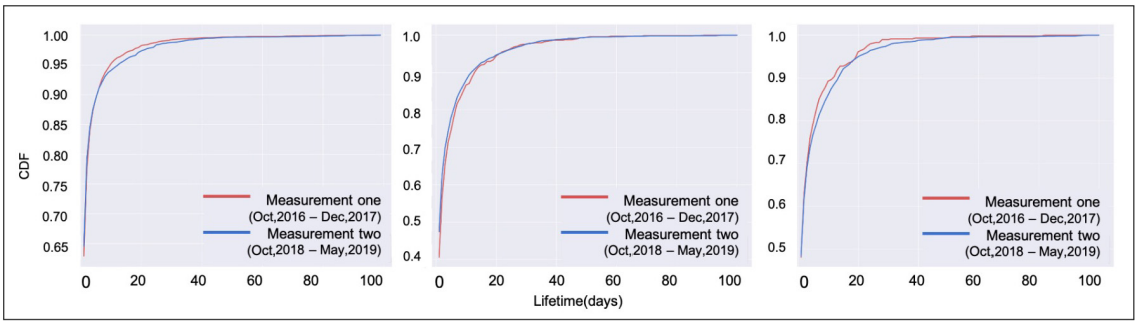


Fig. 4 CDF of malware binary lifetime (left), download server lifetime (middle), and C&C server lifetime (right).

Table 5 Summary of malware binaries, download servers, and C&C servers (CPU Arch. MIPS).

	Measurement one			Measurement two		
Family	**#Binaries**	**#DL**	**#C&C**	**#Binaries**	**#DL**	**#C&C**
Bashlite	894	402	543	161	117	137
Mirai	192	126	41	4,491	2,422	2,176
Tsunami	12	12	12	17	12	10
Total	1,099	540	596	4,669	2,492	2,244

ble 5 shows the detail of these binaries with the number of download (DL) servers and C&C servers. During measurement one, hundreds of Bashlite binaries were captured every day. This may seem that a number of Bashlite botnets are expanding their infection and that the same binaries are continuously downloaded. However, this may not be the case. We estimated the lifetime of these malware binaries and took a deep look into how frequent botnet operators change their infrastructure. We define the lifetime of a binary as the amount of days from the first time it was seen until the last time it is seen. The left graph of **Fig. 4** shows the empirical CDF of the lifetime of malware binaries. About 80% of the binaries were only seen for less than 3 days and 85% of the binaries were only seen for less than 5 days.

This evidences that the lifetime of these binaries were short, indicating that the malware update frequency is high. There were not significant differences among the different malware families, i.e, independently of the malware type the binaries were updated at high frequencies.

Similarly, we estimated the lifetime of the binaries observed in measurement two. The left graph of Fig. 4 shows the empirical CDF of their lifetime. The lifetime of these malware families was also short, which indicates that the malware update frequency has not much changed during measurement one and two. Compared to measurement one, the number of Bashlite binaries decreased while the number of Mirai binaries increased. The trend confirms prior work [10], [24]. As Mirai's source code is based on Bashlite, it seems that botnet operators shifted to Mirai.

The facts that the DL periods of these binaries are very short and that each binary only contains a single C&C domain/IP as shown in previous subsection suggest that these binaries are treated as "disposable". This may be related to the fact that these binaries are non-persistent and can be removed by rebooting the infected devices.

4.2 Infrastructure Server Update Frequency

After seeing that the lifetime of the malware binaries is short, we look into the lifetime of malware download servers and C&C servers. Using the dataset in Section 4.1, we analyzed the number of download servers observed among measurement one and two. Table 5 shows the detail of download (DL) servers during measurement one. Compared to the number of binaries captured, the number of download servers were small. This suggests that different malware binaries are downloaded from the same download server. The middle graph of Fig. 4 shows the empirical CDF of the uptime of download servers. Although, some download servers survive for more than 100 days, we can see that most of the download servers lifetime is short.

We also looked into the number of download servers in measurement two. Compared to the number of binaries captured, the number of download servers were small. The middle graph of Fig. 4 shows the empirical CDF of the uptime of download servers. Although some servers survive a long period, we can see that most of the download servers lifetime is short, which is consistent with prior work [10]. This also indicates that the update frequency has not much changed during the measurements.

We now focus on 1,432 download servers from which we could collect malware binaries for two or more consecutive days. We calculate the frequency of a download server downloading a new binary f_i as below. Where i represents a specific download server; T_i is the uptime of a given download server; and $N_{binaries}(t)$ is the number of new malware binaries downloaded at a certain date t.

$$f_i = \frac{T_i}{\sum_{t=1}^{T_i} N_{binaries}(t)}$$

As the average frequency of each malware family speed, Mirai is 4.8 days. That means each download server updates a new malware binaries in 4.8 days on average. Same for Bashlite is 3.7 days and Tsumani is 0.97 days. On the other hand, out of the 3,032 overall download servers, 1,600 of them (about 53%) were seen within a day. Indicating that more than the half of the download servers were seen only for a short period and were downloading one type of malware binary.

To analyze the lifetime of C&C servers, we used the dataset in Section 4.1. We executed these binaries in a sandbox environment right after a new malware sample was captured by the IoT honeypot. We abstracted connections to C&C servers and as a result, we obtained 2,634 IP addresses and 113 domains as C&C servers. Table 5 shows the detail of C&C servers during measurement one and two. The number of C&C servers are similar to the

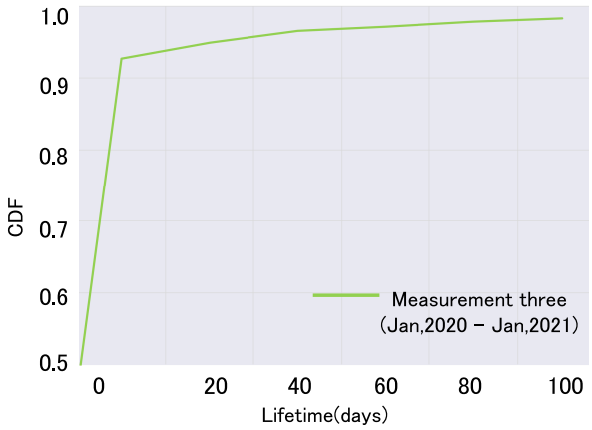


Fig. 5 CDF of C&C server lifetime in measurement three.

number of download servers. In fact, some download servers and C&C servers had the same IP address. We suggest that some botnet operators are using the same machine for both downloading malware binaries and controlling them. We then estimated the lifetime of C&C servers. We define the lifetime of a C&C server as the amount of days from the first time it was seen during the dynamic analysis, until the last time it was seen from another binary's dynamic analysis. In the case of domains, we first obtained all domains from the sandbox analysis. We then abstracted the corresponding IP addresses from DNSDB [25]. At last, we checked if the domain or the corresponding IP addresses were used by the malware binaries during sandbox analysis. The lifetime of this domain is between the first time and last time when the domain/IP address is used by the analyzed binaries. The right graph of Fig. 4 shows the empirical CDF of the uptime of C&C servers. Among both download servers and C&C servers, about 70% of them were active for 5 days or less, and 90% of them were active for 14 days or less. Although some infrastructure servers survive for more than 150 days, most servers have only a short lifetime, which is consistent with prior work [11].

For more long-term analysis, we deployed our IoT honeypot sensor during a relatively recent time period and collected new malware binaries from January 1st, 2020 to January 21st, 2021 (henceforth measurement three). We conducted a 5 minute short-term dynamic analysis for analyzing the lifetime of C&C servers. We executed 4,376 malware binaries that were targeting CPU architecture of MIPS. The analysis was performed after a certain period of time and not immediately after the binary was collected. However, because the sandbox environment does not allow external connections except for domain resolution and most of the captured binaries did not use domains, the influence of this time delay is limited. This is also expressed in the previous study; only a very few samples rely on DNS [12]. As a result, the lifetimes of most of C&C servers were still short. **Figure 5** shows the empirical CDF of the uptime of C&C servers. The lifetimes of about 90% C&C servers were less than 5 days and 93% were less than 14 days. During measurement three, the tendency of short lifetime has not changed but the lifetime of C&C servers seems even shorter.

4.3 C&C Server Connectivity

Besides the lifetimes of malware binaries and their C&C server are short, we analyze the connectivity to their C&C server. We abstracted IP addresses and domain names of their C&C servers by combining both dynamic analysis and static analysis. From the dynamic analysis result in Section 4.2, more than 80% (4,725/5,768) of the binaries tried to connect to their C&C server. We note that only 6% (279/4,725) tried to connect to more than one C&C servers. Most of the malware binaries were directly accessing to their C&C servers using IP addresses while, 10% (475/4,725) tried to connect with domain names. Only 4 binaries used two domains and the rest of the binaries were using a single domain. Furthermore, we randomly chose 1,443 binaries from measurement one and analyzed them using the debugging tool IDA Pro [26]. By checking the disassemble code, we found that these malware binaries only contained a single IP address for connecting. Contrary to traditional Windows malware which often contains multiple C&C server information and uses domains for their robust control, the observed malware binaries did not seem to have such robustness in C&C connection.

Seeing these results, we analyzed malware binaries for a relatively long period and investigated their C&C connection. Due to that the long term malware dynamic analysis occupies the sandbox system, we only executed 50 malware binaries. Out of the binaries, 3 of them tried to connect to its C&C server using domain names but still all of the samples used only one IP address to connect to their C&C server. Most of the binaries lost their C&C connection within 1 or 2 days. The shortest C&C connection was 30 minutes and the longest C&C connection was 142 hours. Even after the C&C connection was lost, all binaries kept trying to connect to their C&C server. The binaries that used a domain for their C&C connection actually looked up the domain only once, right after they were executed. While traditional Windows malware periodically resolves C&C server domains, these IoT malware binaries do not seem to have a robust C&C connection. This may be related to the fact that infections for these malware families are not persistent. Rebooting the infected IoT devices erases the malware process [20].

In measurement three, out of the binaries, more than 80% (3,498/4,376) of them tried to connect to their C&C server. Only 40 (about 0.9%) binaries tried to connect to multiple C&C servers. On the other hand, about 13% (579/4,376) of the binaries tried to connect to their C&C server with domain names. The number of unique IP addresses of C&C servers were 1,989, and the number of unique domains of C&C servers were 129. Therefore, the amount of malware binaries that tried to connect to more than one C&C server decreased (6% in measurements one and two) but the amount of malware binaries that used domain names to connect to their C&C servers increased (10% in measurements one and two). We note that three malware binaries were connecting to Tor nodes. Two binaries were connecting to more than 30 IP addresses toward port 9050/tcp and one binary was connecting to 13 IP addresses towards different tcp ports. This kind of malware binary was also reported in the prior study [12]. We then conducted a long-term dynamic analysis against one binary during March 17, 2021, to March 22, 2021. Although we filtered out-

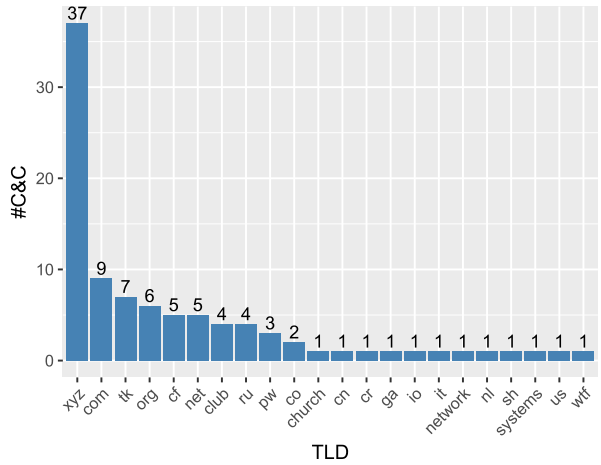


Fig. 6 Amount of C&C per TLD (Oct, 2016 to May, 2019).

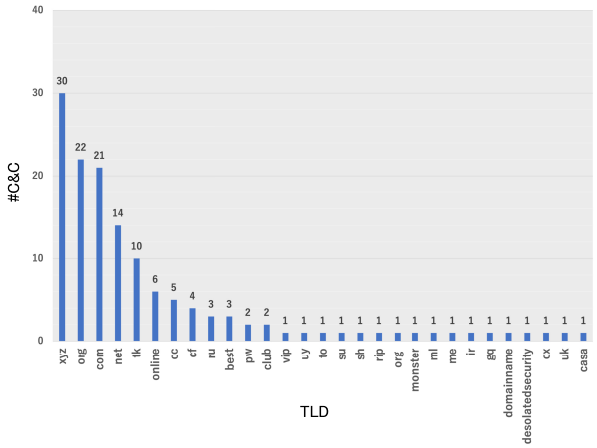


Fig. 7 Amount of C&C per TLD (Jan, 2020 to Jan, 2021).

going packets, we allowed the communication on port 9050/tcp. As a result, connections were established with 136 IP addresses, and communication started with 11 IP addresses. While the shortest connection was a few seconds, others were a relatively long time, 120 to 150 seconds. However, these communications were confirmed only for 1 hour and 30 minutes after the dynamic analysis started. For the IP addresses with which communication was established, 113 communications were established for the most frequent ones, and 18 communications were established for the least ones. Although the binary was collected in August 2020, it was still able to connect to their Tor nodes in March 2021.

4.4 Domain Name of Infrastructure Servers

While the majority of the communication between the infected bots and the C&C servers does not require name resolution, the results of the dynamic analyses showed that a handful binaries did use a few domain names to reach their C&C server. We identified 113 domain names for this purpose – 6 of these corresponded to non-registered domains. We did not see any binary using domains to contact the download servers.

We further investigate the syntax of these domains to gain insights in the distribution of top-level domains (TLDs). **Figure 6** shows the frequency of C&Cs per TLD. Interestingly, 47.8% of these domains are registered in new gTLDs, while 30.1% in ccTLDs and 22.1% in gTLDs. `.xyz` TLD is the largest new gTLD and thus attractive for abuse [27]. Surprisingly, botmasters did not hide the purpose of these domains as the own names used clearly indicated the presence of C&C. That is 62% of the C&C domain names had `cnc`, 3% had `c2`, 2% had `cncc` as subdomain. This evidences the disposability of these domains.

In measurement three, similarly to our old data, the majority of the communication between the infected bots and the C&C servers does not require name resolution. Still, a handful binaries did use a few domain names to reach their C&C servers. Therefore, we further investigated the syntax of these domains. We analyzed TLD of 139 domains used to connect to C&C servers and found that the number of `.xyz` TLD were 30, the number of `.org` TLD were 22, and the number of `.com` TLD were 21. **Figure 7** shows the frequency of C&Cs per TLD. During measurement three, there has been no significant change in the tendency for TLD of C&C servers. However, the number of domains including `cnc` in their names were only 54 (about 39%), and the number of domains including `c2` in their names were only 1 (about 0.7%). This indicates that domains with clear presence of C&C have decreased (62% or 2% in measurement one and two) and that it is becoming more difficult to guess the purpose of using the domains.

4.5 Botnet Clusters

We now concentrate on how binaries, download servers and C&C servers are related to each other. Out of the 50,026 malware binaries collected among measurement one, we abstracted 11,807 binaries that were targeting MIPS and/or MIPSEL CPU architectures. Out of them, we focused on 4,513 binaries that belonged to malware families of Bashlite, Mirai, and Tsunami. Using the dynamic analysis result, we clustered binaries that tried to connect to C&C servers along with their download server and C&C server information. **Figure 8** illustrates the relationship between malware binaries, download servers, and C&C servers. The gray points are Bashlite binaries, the orange are Mirai binaries, and the green are Tsunami binaries. Similarly, the red triangles are the download server IP addresses, the pink triangles are the C&C server IP addresses, and the blue triangles are IP addresses working as both malware download server and C&C server.

In Fig. 8, there is a big group on the top left that includes many Mirai binaries and download servers. This is an artifact of a dummy C&C server in the original Mirai source code that the Mirai author left. So the group indicates that there are many attackers who use the Mirai source code without removing the dummy server, depicted in the center of the group. Excluding the biggest group, we are left with much smaller groups with a single or a few servers. Namely, an IP address is used only for distributing and/or controlling a handful of binaries, forming many disconnected small groups in Fig. 8. Although there is not a clear indicator, we suspect that some of these IP addresses are related to each other and may be used by the same attacker. We will discuss this in Section 4.6.

From Fig. 8, we can also see that there are servers that distribute binaries labelled as different malware families. This indicates that some attackers may be using more than one malware family, which is possible as source codes of these malware are

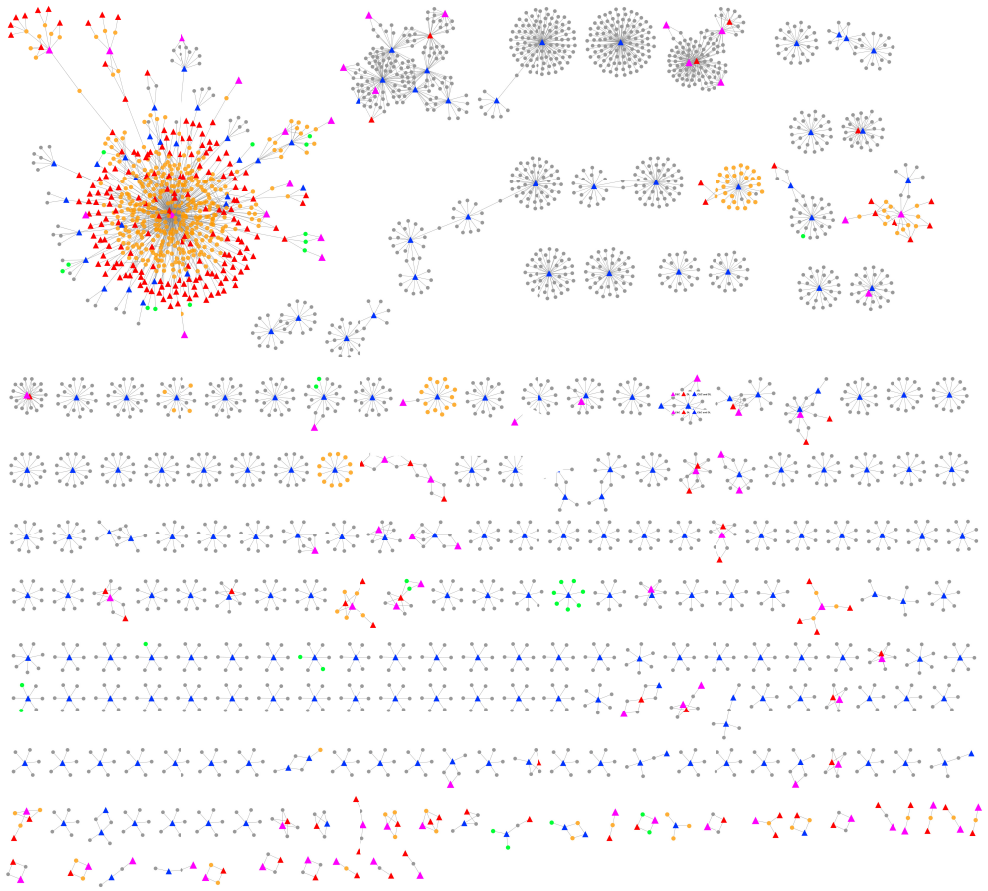


Fig. 8 Relationship of malware binaries, download servers, and C&C servers. Where gray points are bashlite, orange are Mirai, and green are Tsunami binaries. Similarly, for the IP addresses, red triangles are download server, pink triangles are C&C server, and blue triangles are as both malware download server and C&C server (Oct, 2016 to Dec, 2017).

already leaked. Furthermore, we can see that many servers work both as malware download servers and as C&C servers.

New bots are recruited all the time, which execute attacks for a very brief moment and are then abandoned, only to be discovered and compromised again shortly later by other bots. These other bots are then also abandoned. This ongoing cycle basically means that the whole botnet infrastructure is consistently reconstituted from scratch. In combination with their aggressive scanning behavior, it paints a picture of highly disposable botnets.

Finally, we compare malware binaries to see the evolution in their binary code. We chose malware binaries collected from October 19th, 2016 to November 12th, 2016 and abstracted 74 malware binaries which were targeting CPU architecture of MIPS. We first extracted printable characters from the binaries using `string` command. We then manually analyzed each strings using IDA Pro [26] and created Yara rules. These rules are used to find malware binaries that have similarity by matching strings included in the command line, usernames for telnet service, passwords for telnet service, user agents, GetBuild, and other major functions. We grouped these binaries into 14 sets. We then chose one representative binary from each sets and compared their code level similarity using plugin of IDA Pro `diaphora`. We were successful to cluster these binaries into 4 sets. Malware binaries in the same cluster had code level similarity that we assume that the same botnet operator were using these binaries. Three clusters were relatively changing their C&C server. We also found that malware binaries in the same cluster were updating their functionalities. From these results, we suppose that there are botnets that are actively updating their infrastructure.

4.6 Infrastructure Hotspots

Finally, to better understand their relationship, we dig deeper into the Autonomous Systems where most of the download servers and C&C servers are located. We first chose malware binaries which were downloaded from servers in AS 31034, the top AS to have Mirai and Tsunami C&C servers and the second for Bashlite C&C servers during measurement one (Table 4). From these servers, 442 binaries were downloaded. The left graph of **Fig. 9** shows the timeline of these binaries. Each dot is a malware binary. Most of the binaries were seen only briefly at a specific time. Although some binaries were seen for a long period. We also analyzed the timeline of the infrastructure servers. The middle graph of Fig. 9 illustrates the timeline of download servers and the right graph of Fig. 9 illustrates the timeline of C&C servers. Each dot is an infrastructure server. Similar to malware binaries, most of the them were seen only a few times and never observed again. The attackers appear to cycle their infrastructure servers through the IP space of the cloud provider.

We conducted the same analysis for AS23033, another hotspot for download and C&C servers during measurement one. From

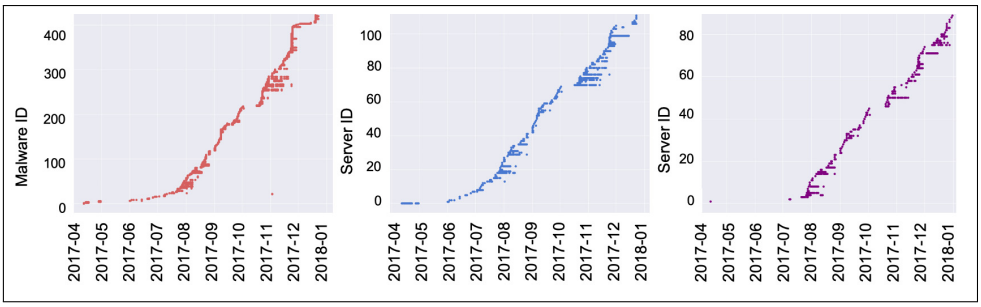


Fig. 9 Malware Binaries (left), download servers (middle), C&C servers (right) seen in AS 31034 (Oct, 2016 to Dec, 2017).

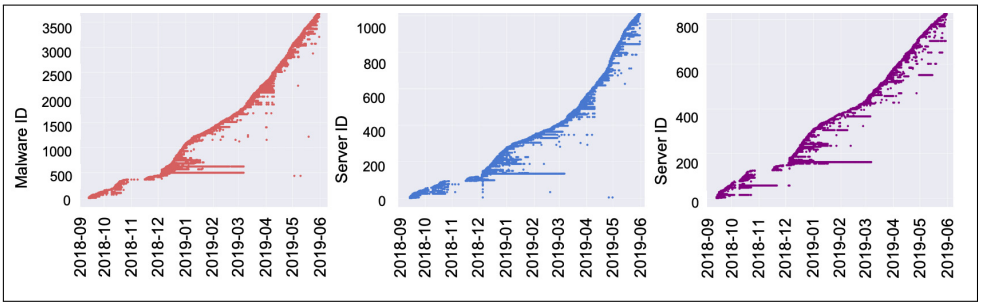


Fig. 10 Malware binaries (left), download servers (middle), C&C servers (right) seen in AS 14061 (Oct, 2018 to May, 2019).

these servers, 580 malware binaries were downloaded. We found the same pattern: the binaries and infrastructure servers were seen briefly and then never observed again.

Although these two ASes were the most observed ones during measurement one, we found that many servers were located in AS 14061 during measurement two. The left graph of **Fig. 10** shows the timeline of these binaries. Even though this analysis was conducted much later, we observe the same patterns. Malware binaries were seen only a few times. We also analyzed the timeline of the infrastructure servers. The middle graph of Fig. 10 illustrates the timeline of download servers and the right graph of Fig. 10 illustrates the timeline of C&C servers. Similar to malware binaries, most of the infrastructure servers were seen only a few times and never observed again. Related to the timeline of the malware binaries, the infrastructure servers were only seen in a certain period, These results point outs that malware binaries that were downloaded from the same AS seems to be shifting and that not much binaries are used at the same time. It can also be said that download servers in the same AS are not used at the same time, equally for C&C servers. We speculate that the attackers might be abusing the IP address allocation practices of cloud providers in order to frequently change the IP addresses of their servers to avoid simple IP-based blacklisting. Namely, the cloud provider of AS14061 provides a virtual server with a static IP address, but the subscriber can change the IP address of the server by simply creating a new copy of the existing server and discarding the old one. This can be done very easily using the snapshot functionality of the cloud provider. It allows attackers to change IP addresses of their C&C server any time they wish at very low cost or no cost. If our speculation is correct, the pattern shown in the timeline of Fig. 10 may indicate that the high count of IP addresses of the C&C servers in the AS is associated with just a few customers, or even a single customer. This implies a small number of threat actors.

As an alternative to bulletproof hosting services [28], which make their servers as takedown-proof as possible, hosting C&C servers in cloud providers seems to be an attractive option [29]. Previous work suggests that we can concentrate mitigation efforts in few networks and penalize infrastructure providers that do not actively take action against these practices [11]. This might be true, but it is not clear that the attackers actually depend on these providers. Since they are all time distributing new binaries with hard-coded C&C IPs, they could basically set up in any cloud provider. The fact that they churn through IPs quickly makes their operations sensitive to how much effort it takes to get a new IP address. This might provide a pressure point to disrupt them.

In measurement three, we picked up 330 malware binaries collected from August 1st, 2020 to August 23rd, 2020 and found 232 IP addresses as their C&C servers. At first, we analyzed AS and network type (e.g., cloud services, access networks, etc.) of IP address of its C&C server. And then, we clarified what kind of network type they were operating at. From this result, the AS of many C&C servers existed in cloud service and there has been no significant change from measurements one and two. Next, we investigated AS 14061 (DigitalOcean) where C&C servers were most concentrated and the next concentrated AS 213371 (ABC Consultancy). The left graph of **Fig. 11** shows the timeline of C&C servers seen in AS 14061 and the right graph of Fig. 11 shows the timeline of C&C servers seen in AS 213371. We found that the same tendency as before was confirmed. In other words, the attackers are still operating their attack infrastructure while frequently changing the address of the C&C server based on the cloud service. In particular, Digital Ocean has been a base for several years.

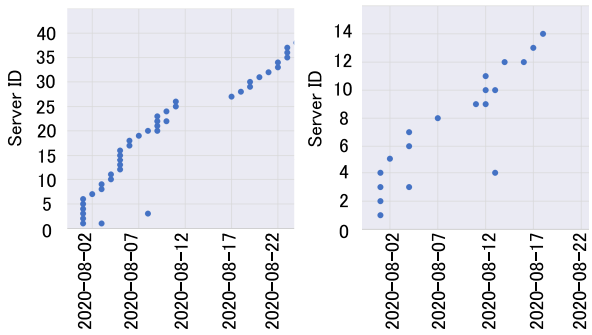


Fig. 11 C&C Servers (left) seen in AS 14061 and C&C servers (right) 213371 (Aug. 1, 2020 to Aug. 23, 2020).

4.7 Summary

To this end, our main insights are as follows:

- The download periods for the malware binaries are very short and each binary typically contains only a single hard-coded IP address or domain. Botmasters that used domains to reach their C&C servers clearly indicated the presence of C&C in their domain names. There was no mechanism for the attackers to migrate the bots to a new C&C server.
- The botnet clusters suggest that operators are not keen about constructing long-lasting botnets with resilient C&C infrastructure, but rather reconstitute their botnets from scratch all the time, driven by aggressive scanning behavior. The picture that emerges is that of highly disposable botnets.
- There are a few ASes where a large portion of the C&C servers are concentrated. Infrastructure servers in such ASes are not in use at the same time and their address space is shifting over time. We speculate that the attackers might be abusing the provisioning mechanisms of cloud providers.

5. Limitations

Our monitoring system presents the common intrinsic limitations of any high-interaction honeypot. Our honeypots present a full operating system for observing skilled attackers that actually target specific vulnerabilities. This allows to filter all the noise created by current scans carried out by researchers and businesses at the expenses of missing some low-skilled attackers.

Beyond the type of interaction that our honeypot requires, the geolocation of the sensors might also decrease the amount of captured attacks. To minimize this impact, we deployed our sensors in multiple regions in Asia and Europe to increase the visibility of the honeypot. However, still there might be some IoT malware that limits its spread to certain regions. Current reports do not seem to indicate the existence of such malware, but they might also suffer from this same limitation.

Similarly, the attack surface of our honeypot is limited to the vulnerable services present in the bare-metal devices. These cover the great majority of well-known infection vectors used by IoT malware. However, most advanced attacks such as the 0-Day RCE Exploit uncovered by the Anglerfish honeypot [30] are not captured by our monitoring infrastructure. While this represents a minority of the current attacks, the infrastructure of the botnets behind these advanced attacks might differ from the botnet families presented in this paper.

Finally, the monitoring period of our infrastructure is limited to 36 months. We believe this period is enough to obtain a comprehensive picture of the infrastructure used by these botnets. However, the rapid evolution of IoT malware requires continuous monitoring to capture the (mis) use of the infrastructure.

6. Related Work

One area of research has focused on the propagation of IoT malware, as observed via honeypots. IoTPOT [18] observed telnet-based attacks into three typical steps and show how attackers intrude, infect, and control the target device. ThingPot [31] found that the attackers seemed to start scanning to look for openings, followed by a more targeted and specific attack via brute force or fuzzing. SIPHON [32] analyzed traffic and attacks across different geo-locations which received significantly different amounts of connections and traffic. X-POT [33] adapted responses collected from the host through Internet-wide scan and observed attacks targeting various IoT devices with integrity. Our study also used a honeypot system to monitor these attacks. Furthermore, we implemented a high interaction honeypot connected with a dynamic analysis system to see how malware binaries, download servers and C&C servers are related to each other to form the backbone of IoT botnets.

Another area of research has focused on analyzing the aspects of IoT botnets. Antonakakis et al. [24] presented the growth, composition and evolution of Mirai botnet. They identified 33 C&C clusters that shared no infrastructure and estimated their relative size. Some botnets were upgrading from IP-based to domain-based C&C connections to avoid detection. At least 17% of the C&C domains were expired and re-registered before being used. Vervier et al. [10] combined low and high interaction honeypots to take a look at how IoT device are compromised. They investigated several IoT malware families and described that most binaries were only distributed for a single day. Moreover, they discovered that malware download servers IP addresses only appeared for a short time period and then disposed. Alrawi et al. [12] studied the lifecycle of IoT malware using a large-scale dataset. They examined more than 138 k malware samples and stated that IoT malware rely mostly on hard-coded IP addresses for C&C call-back. Bastos et al. [11] analyzed Bashlite and Mirai's C&C servers using honeypots, dynamic and static analysis, and active clients. It has been shown that the lifetime of C&C servers were short, most of them were seen only for a few days, and that 84% of them were hosted in cloud providers. Artur et al. [13] deployed low interaction honeypots and monitors that connect to C&C servers. They found that malware download servers and C&C servers of Bashlite and Mirai were concentrated in few ASes, mainly hosted on cloud providers. Stephen et al. [16] measured the size of Hajime botnet and revealed that they had a particularly heavy concentration in a small number of countries, 52.5% in Brazil. They identified that Hajime had a high churn among IoT devices that, mostly their lifetime were less than 5 hours. Our research builds on these works and presents a longer-term study. First, we add the analysis of how malware binaries connect to their C&C servers and how attackers update their binaries or C&C server information. Second, we investigate

how binaries, download servers and C&C servers are related to each other. Finally, we look into how they evolve over time and how attackers source their IP addresses.

7. Conclusion

We designed a honeypot system for monitoring the evolution and characteristics of IoT botnet infrastructure. Over the course of 36 months , we collected 64,260 IoT malware binaries and data on 35,494 download servers and 4,736 C&C servers. We then analyzed the botnet infrastructure of Bashlite, Mirai, and Tsunami from October 2016 to December 2017 and October 2018 to May 2019. Malware binaries were distributed only briefly, typically for less than five days. We analyzed their connectivity to the C&C and revealed that most of these binaries only contained a single hard-coded C&C IP address or, in 10% of the cases, a domain. The information for the download servers is also hard-coded into the binary. The download and C&C servers were located in cloud providers. The servers were only available for a short time, typically less than 5 days. The long-term dynamic analysis showed that the C&C server made no attempts to update the binary with a newer version or with refreshed information on C&C or download servers. Once the binary no longer received a response from the server, it had no way of re-establishing a connection to the botnet. There were a few ASes where a large portion of the botnet infrastructure was concentrated. These C&C servers were typically used sequentially, rather than in parallel. So the perceived concentration is not so much the density of C&C, but one or a few servers that are moving through the address space of the provider. Although malware binaries that use domain names to connect to their C&C servers increased in 2020, the C&C servers themselves were still short-lived and the tendency has not changed for years. Moreover, the same cloud providers were abused and yet had been a base for several years.

In sum, our analysis reveals a very clear pattern: most bots are abandoned by the attackers within a few days. Even though the abandoned bots do continue to scan aggressively for vulnerable hosts, after a few days they cannot capture new hosts for the botnet anymore, since the download server where they attempt to get the binary is no longer available at the hard-coded IP address. All in all, attackers treat IoT botnets as wholly disposable. The botnet population and the associated C&C and download infrastructure is reconstituted again and again. This pattern has been going on for five years now and shows no sign of changing.

Compared to Windows botnets, it is easy to see the architecture of IoT botnets as primitive, or even amateuristic, but that seems to overlook the fact that their disposable nature makes them very resistant to blacklisting or takedown of C&C servers. As they are being reconstituted again and again, for more than three years now, it seems that the only effective countermeasure is to either identify and apprehend the criminals, or to remediate the enormous population of vulnerable devices. Both of these tasks are not impossible—criminals have been arrested [34] and consumers have been remediating their devices [20]. This progress has not been enough, however, to really change the landscape of IoT botnets.

Acknowledgments This research was partly conducted under a contract of “Research and development on IoT malware removal / make it non-functional technologies for effective use of the radio spectrum” among “Research and Development for Expansion of Radio Wave Resources (JPJ000254)”, which was supported by the Ministry of Internal Affairs and Communications, Japan. This work was also partly financed by the Dutch Research Council (NWO), under project number 628.001.033, and by SIDN, the .nl registry, under the MINIONS-NL project grant.

References

[1] ZDNet: Hacker leaks passwords for more than 500,000 servers, routers, and IoT devices (2019), available from ⟨https://www.zdnet.com/article/hacker-leaks-passwords-for-more-than-500000-servers-routers-and-iot-devices/⟩.

[2] Pour, M.S., Mangino, A., Friday, K., Rathbun, M., Bou-Harb, E., Iqbal, F., Shaban, K. and Erradi, A.: Data-driven Curation, Learning and Analysis for Inferring Evolving IoT Botnets in the Wild, *14th International Conference on Availability, Reliability and Security* (*ARES2019*), ACM (2019).

[3] Spring, T., Carpenter, K. and Mimoso, M.: BASHLITE family of Malware Infects 1 Million IoT devices (2016), available from ⟨https://threatpost.com/bashlite-family-of-malware-infects-1-million-iot-devices/120230/⟩.

[4] Krebs on Security: Who is Anna-Senpai, the Mirai Worm Author? (2017), available from ⟨https://krebsonsecurity.com/2017/01/who-is-anna-senpai-the-mirai-worm-author/⟩.

[5] Barnett, R.: New Tsunami/Kaiten Variant: Propagation Status - Akamai Security Intelligence and Threat Research Blog (2018), available from ⟨https://blogs.akamai.com/sitr/2018/09/new-tsunamikaiten-variant-propagation-status.html⟩.

[6] Donno, M.D., Dragoni, N., Giaretta, A. and Spognardi, A.: Analysis of DDoS-Capable IoT Malwares, *Federated Conference on Computer Science and Information Systems* (2017).

[7] Bijmans, H.L., Booij, T.M. and Doerr, C.: Just the Tip of the Iceberg: Internet-Scale Exploitation of Routers for Cryptojacking, *ACM SIGSAC Conference on Computer and Communications Security*, pp.449–464 (2019).

[8] Fortinet: DDoS-for-Hire Service Powered by Bushido Botnet (2018), available from ⟨https://www.fortinet.com/blog/threat-research/ddos-for-hire-service-powered-by-bushido-botnet-.html⟩.

[9] 360 Netlab Blog: Art of Steal: Satori Variant is Robbing ETH BitCoin by Replacing Wallet Address (2018), available from ⟨https://blog.netlab.360.com/art-of-steal-satori-variant-is-robbing-eth-bitcoin-by-replacing-wallet-address-en/⟩.

[10] Vervier, P.-A. and Shen, Y.: Before Toasters Rise Up: A View Into the EmergingIoT Threat Landscape, *Proc. 21st International Symposium on Research in Attacks, Intrusions, and Defenses* (*RAID 2018*), pp.556–576, Springer (2018).

[11] Bastos, G., Marzano, A., Fonseca, O., Fazzion, E., Hoepers, C., Steding-Jessen, K., Chaves, M.H., Cunha, I., Guedes, D. and Meira, W.: Identifying and Characterizing Bashlite and Mirai C2C Servers, *ISCC 2019*, Barcelona, Spain (2019).

[12] Omar, A., Charles, L., Kevin, V., Ryan, C., Kevin, S., Fabian, M. and Manos, A.: The Circle Of Life: A Large-Scale Study of The IoT Malware Lifecycle, *30th USENIX Security Symposium* (*USENIX Security 21*), USENIX Association (2021).

[13] Marzano, A., Alexander, D., Fonseca, O., Fazzion, E., Hoepers, C., Steding-Jessen, K., Chaves, M.H.P.C., Cunh, I., Guedes, D. and Meira, W.: The Evolution of Bashlit and Mirai IoT Botnet, *IEEE Symposium on Computers and Communications* (*ISCC 2018*) (2018).

[14] Edwards, S. and Profetis, I.: Hajime: Analysis of a decentralized internet worm for IoT devices, *Rapidity Networks*, Vol.16 (2016).

[15] van der Wiel, J., Diaz, V., Namestnikov, Y. and Konstantin, Z.: Hajime, the mysterious evolving botnet (2017), available from ⟨https://securelist.com/hajime-the-mysterious-evolving-botnet/78160/⟩.

[16] Herwig, S., Harvey, K., Hughey, G., Roberts, R. and Levin, D.: Measurement and Analysis of Hajime, a Peer-to-peer IoT Botnet, *The Network and Distributed System Security Symposium* (*NDSS 2019*) (2019).

[17] Tanabe, R., Tamai, T., Fujita, A., Isawa, R., Yoshioka, K., Matsumoto, T., Gañán, C. and Van Eeten, M.: Disposable Botnets: Examing the Anatomy of IoT Botnet Infrastructure, *15th International Conference on Availability, Reliability and Security* (*ARES2020*), pp.7:1–7:10, ACM (2020).

[18] Pa, Y.M.P., Suzuki, S., Yoshioka, K., Matsumoto, T., Kasama, T. and

Rossow, C.: IoTPOT: Analysing the Rise of IoT Compromises, *9th USENIX Workshop on Offensive Technologies* (*WOOT 15*), USENIX Association (2015).
[19] Bernstein, J. and Spets, T.: CPE WAN management protocol, *DSL Forum, Technical Report TR-069* (2004).
[20] Cetin, O., Ganan, C.H., Altena, L., Kasama, T., Inoue, D., Tamiya, K., Tie, Y., Yoshioka, K. and van Eeten, M.: Cleaning Up the Internet of Evil Things: Real-World Evidence on ISP and Consumer Efforts to Remove Mirai, *NDSS 2019* (2019).
[21] Daniloff, I.: Doctor Web (2019), available from ⟨https://www.drweb.com⟩.
[22] Maxmind: GeoIP2 Databases (2019), available from ⟨https://www.maxmind.com/⟩.
[23] IP2Location.com: IP2Location (2018), available from ⟨https://www.ip2location.com/⟩.
[24] Manos, A., Tim, A., Michael, B., Matt, B., Elie, B., Jaime, C., Zakir, D., J. Alex, H., Luca, I., Michalis, K., Deepak, K., Chaz, L., Zane, M., Joshua, M., Damian, M., Chad, S., Nick, S., Kurt, T. and Yi, Z.: Understanding the Mirai Botnet, *26th USENIX Security Symposium* (*USENIX Security 17*), Vancouver, BC, pp.1093–1110, USENIX Association (2017).
[25] Farsight Security: DNSDB (2018), available from ⟨https://www.dnsdb.info/⟩.
[26] Hex-Rays: IDA Pro (2019), available from ⟨https://www.hex-rays.com/products/ida/⟩.
[27] Korczynski, M., Wullink, M., Tajalizadehkhoob, S., Moura, G.C. and Hesselman, C.: Statistical Analysis of DNS Abuse in gTLDs Final Report, Technical report, Technical Report (2017), available from ⟨https://www.icann.org/en/system/files/files/sadag-final-09aug17-en.pdf⟩.
[28] Goncharov, M.: Criminal hideouts for lease: Bulletproof hosting services (2015), available from ⟨https://www.trendmicro.no/media/wp/wp-criminal-hideouts-for-lease-en.pdf⟩.
[29] Khattak, S., Ramay, N.R., Khan, K.R., Syed, A.A. and Khayam, S.A.: A Taxonomy of Botnet Behavior, Detection and Defense, *IEEE Communications Surveys and Tutorials* (2013).
[30] Ye, G.: Hunting advanced IoT malware, *The 22nd International AVAR Cybersecurity Conference*, available from ⟨https://www.avar2019.org/agenda/day-1/hunting-advanced-iot-malware⟩ (2019).
[31] Wang, M., Santillan, J. and Kuipers, F.: ThingPot: An interactive IoT honeypot, available from ⟨https://arxiv.org/abs/1807.04114⟩ (2018).
[32] Guarnizo, J., Tambe, A., Bhunia, S.S., Ochoa, M., Tippenhauer, N., Shabtai, A. and Elovici, Y.: Siphon: Towards scalable high-interaction physical honeypots, *CPSS 2019* (2019).
[33] Seiya, K., Rui, T., Katsunari, Y. and Tsutomu, M.: Adaptive Observation of Emerging Cyber Attacks targeting Various IoT Devices, *2021 IFIP/IEEE International Symposium on Integrated Network Management* (*IM 2021*), IEEE (2021).
[34] Krebs on Security: Mirai IoT Botnet Co-Authors Plead Guilty (2017), available from ⟨https://krebsonsecurity.com/2017/12/mirai-iot-botnet-co-authors-plead-guilty/⟩.

Rui Tanabe received his Ph.D. in information sciences from Yokohama National University in 2017. After working at Yokohama National University as a researcher, He is currently working as a project associate professor at Yokohama National University. His research interests include information security and network security. He received the Yamashita Memorial Research Award from IPSJ in 2017.

Tsuyufumi Watanabe received his M.E. in Musashi Institute of Technology in 1999. He is currently working as General Manager of Security Management Department at FUJISOFT Incorporated. His research interests include information security and network security. He is also a Ph.D. student at Yokohama National University.

Akira Fujita received B.A., M.Sc. and Ph.D. in information sciences from Yokohama National University in 2008, 2009 and 2012 respectively. After working as a project assistant professor at Yokohama National University and National Institute of Informatics, he is a senior researcher at National Institute of Information and Communications Technology. His research interests include network security, natural language processing and cognitive science.

Ryoichi Isawa received his B.E. and M.E. degrees from the University of Tokushima, Japan, in 2004 and 2006, respectively. He received his Ph.D. degree from Kobe University, Japan, in 2012. He is currently a senior researcher at the National Institute of Information and Communications Technology (NICT), Japan. His current research interests include malware analysis, hardware security, and information security. He is a member of the IEEE and IEICE.

Carlos Gañán is an associate professor in the Economics of Cybersecurity group at TU Delft. His research focuses on the cybersecurity of the Internet of Things, Smart Cities, and Data Science and explores the security requirements of connected products and services for a sustainable future. In 2012, he completed his Ph.D. in the field of information security for vehicular ad-hoc networks. Previously, in 2008 he completed a MSc on Telecommunications writing a thesis on the safety and security of wireless sensor networks at Philips Laboratories in Aachen. After that, in 2010 he received an M.Sc. in Telematics during which he studied the secure transmission of video streaming for mobile ad-hoc networks. In the past, he was part of the Information Security Group, with the Department of Telematics Engineering at UPC, Barcelona. He also holds a Diploma in Business Studies and a Degree in Administration and Business Management from the Open University of Catalonia.

Michel van Eeten is a professor of cyber-security at Delft University of Technology. His team analyses large-scale Internet measurement and incident data to identify how the markets for Internet services deal with security risks. He has conducted empirical studies for the ITU, OECD, European Commission and the Dutch government on the economics of malware, the impact of cybercrime and the role of ISPs in mitigating botnets and bad hosting. He is also a member of the Dutch Cyber Security Council.

Katsunari Yoshioka is an Associate Professor at Yokohama National University since 2011. Before that, he was a researcher at National Institute of Information and Communications Technology, Japan. His research interests cover wide area of system security and network security including malware analysis and IoT security. He received the commendation for science and technology by the minister of MEXT, Japan in 2009, the award for contribution to Industry-Academia-Government Collaboration by the minister of MIC, Japan in 2016, and the Culture of Information Security Award in 2017.

Tsutomu Matsumoto is a professor of the Faculty of Environment and Information Sciences, Yokohama National University, and directing the Research Unit for Information and Physical Security at the Institute of Advanced Sciences. He also serves as the Director of the Cyber Physical Security Research Center (CPSEC) at the National Institute of Advanced Industrial Science and Technology (AIST). Starting from Cryptography in the early ’80s, he has opened up the field of security measuring for logical and physical security mechanisms. He received a Doctor of Engineering degree from the University of Tokyo in 1986. Currently, he is interested in research and education of Embedded Security Systems such as IoT Devices, Cryptographic Hardware, In-vehicle Networks, Instrumentation and Control Security, Tamper Resistance, Biometrics, Artifact metrics, and Countermeasure against Cyber-Physical Attacks. He serves as the chair of the Japanese National Body for ISO/TC68 (Financial Services) and the Cryptography Research and Evaluation Committees (CRYPTREC) and as an associate member of the Science Council of Japan (SCJ). He was a director of the International Association for Cryptologic Research (IACR) and the chair of the IEICE Technical Committees on Information Security, Biometrics, and Hardware Security. He received the IEICE Achievement Award, the DoCoMo Mobile Science Award, the Culture of Information Security Award, the MEXT Prize for Science and Technology, and the Fuji Sankei Business Eye Award.