Examination of Investigation Method of Malware Spreading State

Anh Son Pham(✉) and Yasuhiro Nakamura

Cyber Security, Graduate School of Science and Engineering,
National Defense Academy, 1-10-20, Hashirimizu, Yokosuka, Japan
{em59043,yas}@nda.ac.jp
http://www.nda.ac.jp/cc/gsse, http://www.nda.ac.jp/~yas/

Abstract. In recent years, there have been many studies and reports on malware infection activities. These studies and reports focused on the number of source IP addresses with malware characteristics. By measuring the number of source IP addresses per day, we can determine the status of malware infection activity. The results were used to alert the public. However, such studies are unable to distinguish between source IP addresses that appear on multiple days and newly appearing source IP addresses, and the circumstances under which malware infections spread or shrink are unknown. Here, we measured the number of newly appearing source IP addresses over a long period of time to reveal the spread or shrinking of the malware.

Keywords: Malware spreading · Malware infection · Mirai · Hajime

1 Introduction

As of 2011, it was estimated that there would be about 50 billion IoT devices in 2020 [1]. On the other hand, it is known that embedded network devices such as IoT devices have many vulnerabilities [2–4]. In recent years, malware that uses these vulnerabilities to infect IoT devices, build overlay networks, and perform DoS attacks has become widespread. For example, in Sep. 2016, a large-scale and destructive DDoS attack was carried out by the malware Mirai. The appearance of botnets that enable large-scale attacks attracted attention. In the United States in 2019, many medical and educational institutions suffered about $7.5billion from the ransomware attack. Data from the 2020 Ransomware Resiliency Report also shows that 35% of organizations suffered between $1million and $5million [5].

In order to prevent such attacks, it is very important to analyze the behavior of malware and monitor the infection status. In particular, research on the infection activity of worm-type malware targeting IoT devices that spread destructively and actively, such as Mirai, is drawing attention. Darknet observation and honey port installation are effective methods for grasping the attack activity of

N.-S. Vo et al. (Eds.): INISCOM 2021, LNICST 379, pp. 128–141, 2021.
https://doi.org/10.1007/978-3-030-77424-0_11

these malwares. However, many general observation reports count the number of unique daily IP addresses that arrive at the sensor. The method can capture the activity of the worm that day, observing the source addresses for each day can be used to investigate the increase or decrease in infection activity, but it cannot distinguish between source addresses that appeared on multiple days and newly appearing source IP addresses, and the actual spread of infection is unknown. Therefore, this paper decided to investigate the infection status of worm-type malware. Using a specific day as the observation starting point, we measure the number of newly appearing source IP addresses for each day since that day. The number of newly appearing source IP addresses per day from that day is measured, and the increase or decrease in the number of infections can be determined by comparing it to the previous and following days. Furthermore, by measuring over a long period of time, it is possible to clarify the expansion and contraction of malware infections.

2 Related Works

2.1 Definitions

Each term is defined as follows.

(1) Darknet
A darknet is an IPv4 address space advertised in BGP, and is a set of addresses to which devices are not connected.
(2) Darknet Observation
Darknet observation refers to the acquisition and accumulation of packets that arrive at an address to which a device is not connected.
(3) The number of Sender Address
The number of source addresses in total number of unique addresses obtained by extracting malware-characteristics from all packets observed during the day and excluding duplicate source addresses.
(4) The number of new source address
The number of source addresses in total number of newly appearing source addresses in D days.

2.2 Related Works

The results of observing the activity of multiple malware that infects IoT devices have been reported in many reports. NICT reported the observation result of the number of packets that arrived on the darknet [6]. This report indicates the days when infectious activity was observed based on the communication characteristics of ransomware such as WannaCry, Petya [7] and BadRabbit [8]. In addition, the date and time when the infection activity of the IoT malware Mirai variant and Hajime was observed and the country statistics of the sender are reported. Also, NICT extracted observational data featuring the malware Mirai and Hajime and reported an increase or decrease in the number of source

IP addresses on a daily basis [9]. As a result, it was found that the number of Hajime infections was on the order of several times the number of Mirai infections. This report also clarified the scale of infection of Hajime and Mirai and the date when the infection activity was activated.

IIJ-SECT Security Report 2018 [10] investigated the infection activity of Mirai, qBot, and Hajime using packets that arrived at Honey Port, and measured changes in the number of IP addresses of these sources. The report clarified the tendency of malware infection activity and showed the number of source IP addresses for each port. In addition, in IIJ-SECT Security Report 2019 [11], it investigated the activity status of Mirai, Hajime, and qBot based on the packet data received at a honeypot. It measured the activity status and infection scale of Mirai variants, and indicated the relationship with the case where the DDoS attack using moobot in September 2019 caused damage to the services of Wikipedia, Twitch, and Blizzard.

These reports focus on the source IP addresses of packets that have the communication characteristics of malware, and grasp the activity of malware infection activity by finding the unique number of source IP addresses for each day. However, with this method, it is not possible to distinguish between a source address that appears on multiple days and a newly appearing source address, so it is not possible to determine the daily increase or decrease of infected addresses. Therefore, in this study, we propose a method to identify a newly infected source address and measure the increase or decrease of infection by finding a unique source address over the entire observation period.

2.3 Communication Features of Hajime and Mirai

Worm-type malware, such as Mirai and Hajime, may have unique characteristics in their scan packets. In this section, we determine that an incoming packet is a connection request from Mirai or Hajime based on the following already known features.

Features of Mirai
The destination port is 23 or 2323, the sequence number and destination address of the TCP packets are the same [9].

Features of Hajime
A lot of research has been done on Hajime, and the communication characteristics of Hajime have been divided into two main categories and studied. In this study, we take these two features and consider them as separate malware:

Hajime - Sequence Number: The upper or lower 16 nits of the sequence number of TCP packets will be zero [12].

Hajime - Window size: The window size of TCP packets is fixed at 14600 [9].

3 Proposal Method

This study analyzes the data that allowed us to observe unauthorized communication in the darknet. The analysis method is divided into five major steps, as shown in Fig. 1.

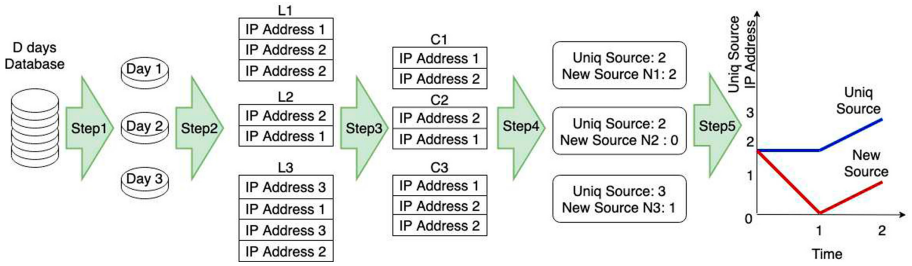

Fig. 1. Proposal method.

Step1: Divide the whole of the D-day's data observed from the darknet into one day's worth of data.

Step2: Extract the source IP addresses of the target packets with malware communication characteristics from the observed data for one day and create the source list Li for day i.

Step3: Use the source list Li created in step 2 to create a unique source list Ci for day i.

Step4: Find the number of new appearing senders Ni from the Ci list for day D.

Step5: Calculate the number of IP addresses in the Ci and Ni lists and draw a graph.

In the resulting graph, the blue line is the daily number of source addresses based on the existing method of NICT, and the red line is the number of newly appearing source addresses based on the proposed method of this study.

4 Experiment

4.1 Experiment Data

This study utilizes data from observations of unauthorized communication in the darknet. The dataset is a darknet observation dataset provided by NICT. The observation period is from 01/01/2016 to 12/31/2018. In addition, the target of this study is the IoT malware Mirai and Hajime, which have become active in recent years. We measure the number of daily source addresses of packets with Mirai and Hajime feature communications and the number of newly appearing

source addresses. In this study, we use darknet observation data provided by NICT.

4.2 Experimental Results

Hajime

Results Based on the Characteristics of the Sequence Numbers

Figure 2 shows the infection activity from 01/01/2016 to 12/31/2018 and the measurement started on 01/01/2016. According to the results, it was confirmed that Hajime was presented even before 2016 and in Oct. 2016 the infection activity started to increase. At this time, the spread of Hajime infection increased along with the increase in infectious activity. The peak of spread of the infection was on 01/12/2017, when approximately 140,000 newly appearing source IP addresses of transmission were identified. It has gradually shrunk until August 2017 but infection activity began to spike in August 2017 and continued until April 2018. On the other hand, Fig. 3 shows that the number of newly appearing source IP addresses decreased and the infections began to shrink. After April 2018, the number of sender addresses became almost non-existent, and it could be confirmed that the number of source IP addresses did not increase again. Figure 3 also shows that the largest number of newly appearing source IP addresses per day in 2018 was about 170,000.

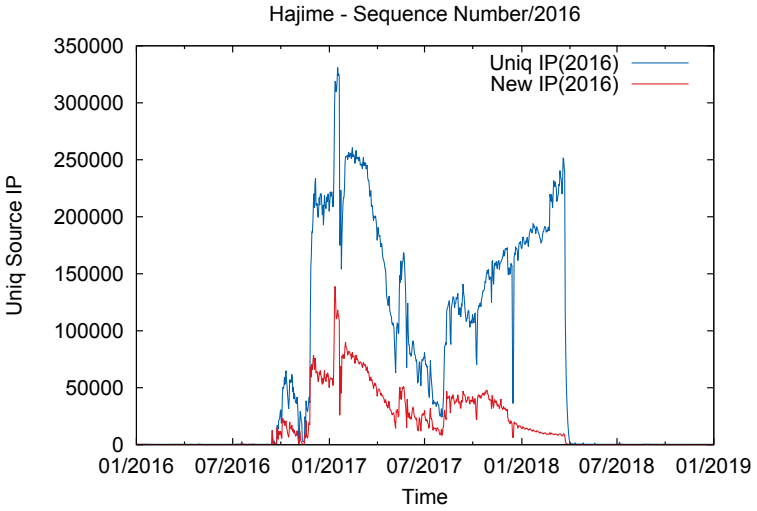

Fig. 2. Spreading of Hajime - Sequence Number from 01/01/2016 to 12/31/2018.

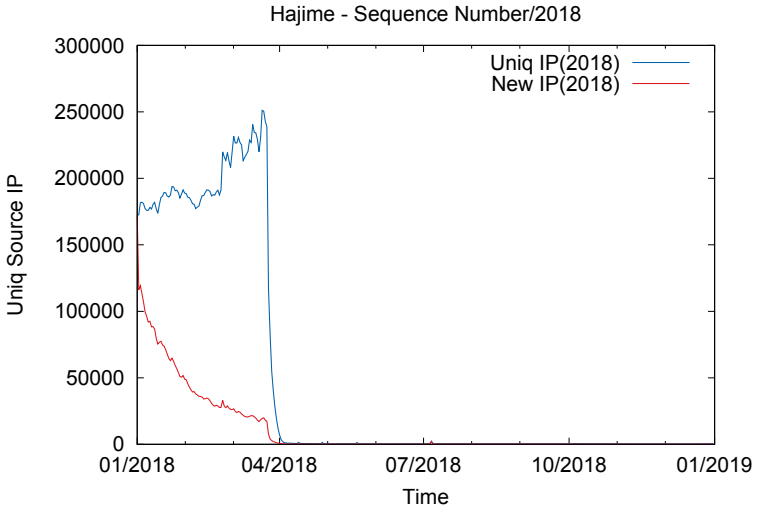

Fig. 3. Spreading of Hajime - Sequence Number in 2018.

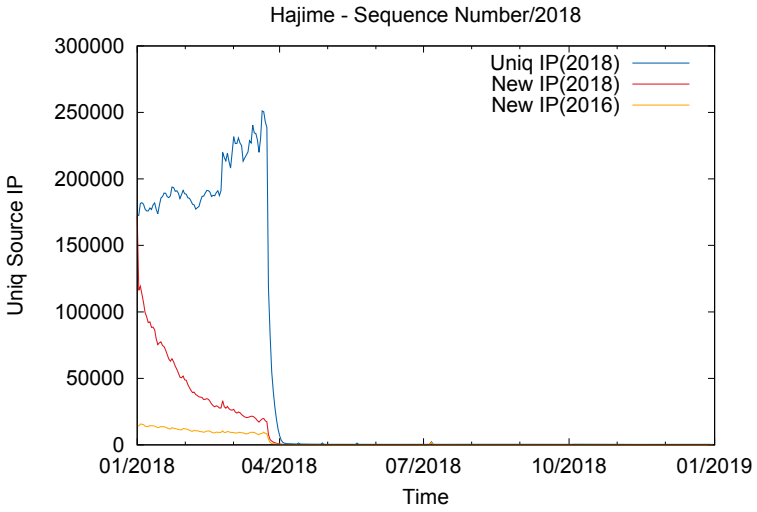

Fig. 4. Spreading of Hajime - Sequence Number in 2018 and newly source IP line 2016.

Figure 4 shows the trend of new source addresses of transmission, with 01/01/2016 as the measurement starting point, introduced into Fig. 3. New IP(2016) is a transition line for the number of newly appearing source addresses for 2018 with starting point on 01/01/2016. We can see that the actual number of new source IP addresses is a small fraction of the total number of source IP addresses that participated in the infection activity. Most of the senders that

engaged in infectious activity in the first half of 2018 were those that had also appeared before 2018. We also find that the largest number of newly appearing source IP addresses per day was about 15,000, not the 170,000 in Fig. 3.

Results Based on the Characteristics of the Window Size
The trend in the number of source addresses of packets with the characteristics of window size is shown in Fig. 5 below. The observation period is 01/01/2016–12/31/2018, and the observation start point is set to 01/01/2016. Hajime was confirmed to have been presented even before 2016, it was active in just a few host units until May 2016. From May 2016 onwards, Hajime's infection has been spreading along with a gradual increase in infection activity. The peak of the spread was on 1/12/2017. Approximately 140,000 hosts could be identified as newly appearing source addresses. Since then, the infections have been gradually decreasing until August 2017.

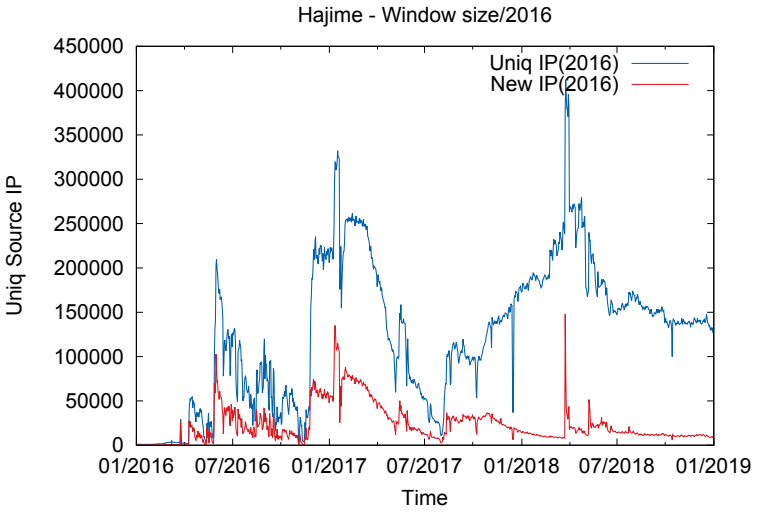

Fig. 5. Spreading of Hajime - Window size from 01/01/2016 to 12/31/2018.

From August 2017 to the end of March 2018, it appeared that the infection was active from August 2017 to the end of March 2018, and a large scale of infection activity took place in April. On the other hand, the number of new source addresses decreased during this period, and the scale of infection decreased. Figure 6 shows that Hajime's infection expanded rapidly on 3/25/2018. More than 148,000 newly appearing source IP addresses were identified, exceeding the peak of 1/12/2017. Since then, the infection has shrunk a bit, but the infection activity was still maintained.

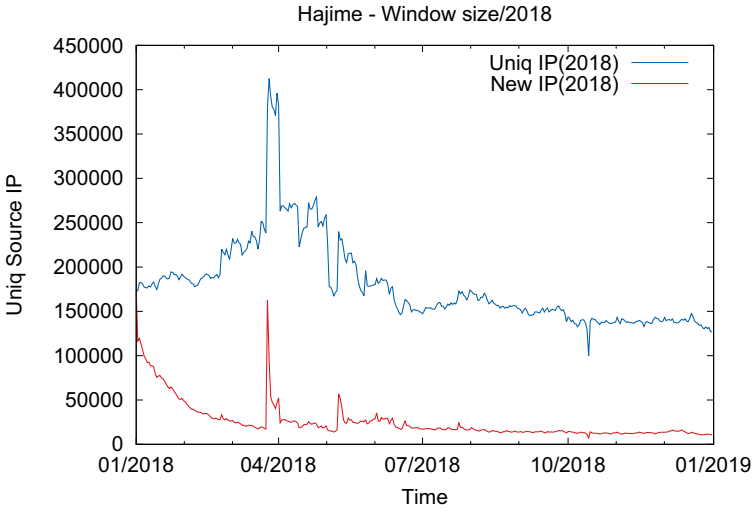

Fig. 6. Spreading of Hajime - Window size in 2018.

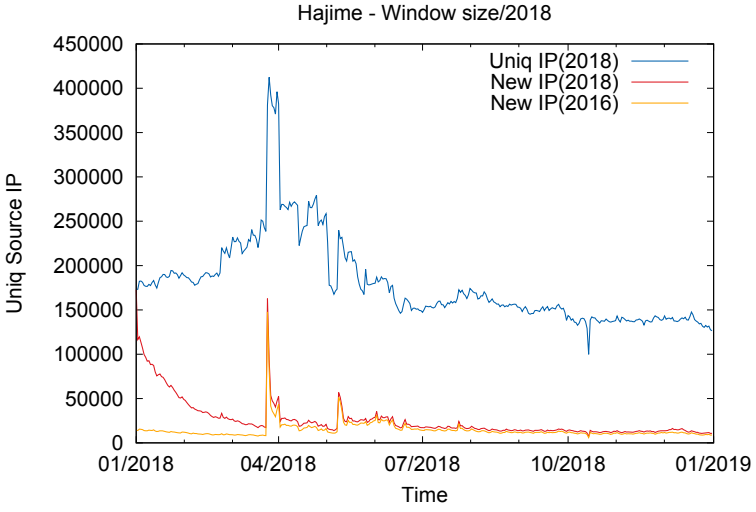

Fig. 7. Spreading of Hajime - Window size in 2018 and newly source IP line 2016.

To get a more accurate picture of the number of new source addresses, Fig. 6 inserts trend of newly appearing source IP addresses starting from 1/1/2016 and Fig. 7 shows that the error in the number of new source addresses up to April was very large. New IP(2016) is a transition line for the number of newly appearing source addresses for 2018 with starting point on 01/01/2016. We can also see that most of the source IP addresses of infection activity in 2018 also appeared before 2018.

Comparison of Results by Window Size and Sequence Number
While the number of source IP addresses having the feature of Window size began to increase from May of 2016, Fig. 8 showed that the number of source addresses with the TCP sequence number feature is from October 2016. The number of source addresses with the TCP sequence number feature from October was the same as the number of source addresses with the window size feature. The number of source addresses with the TCP sequence number feature decreased sharply from the end of March 2018, while the number of source addresses with the window size feature increased rapidly.

In addition, although newly appearing source addresses with characteristics of Window size from Fig. 9, Hajime infections began to spread from March 2016. The results of the study on the characteristics of TCP sequence numbers confirmed that it was from October 2016. From December 2016 to the end of March 2018, we confirmed that the results were the same as those studied in the Window size feature. In April 2018, fewer source addresses have the characteristics of the TCP sequence number, and the infection has shrunk, but a sharp increase in the number of source addresses that had the characteristics of Window size could confirm that the infection had spread rapidly.

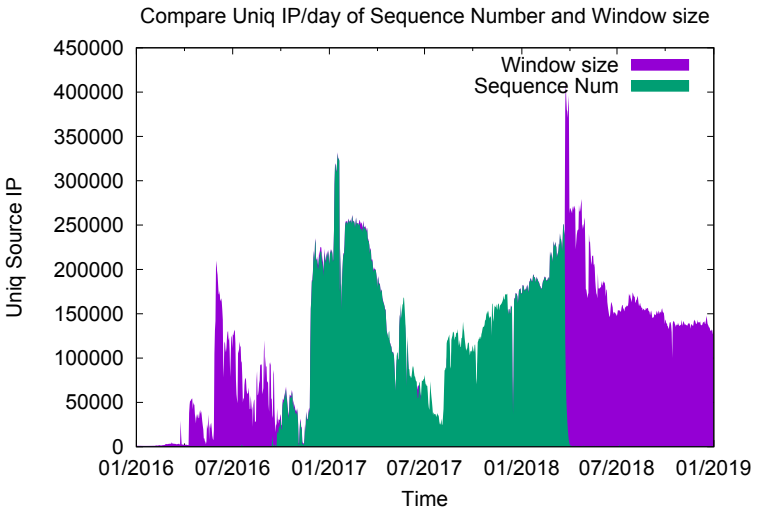

Fig. 8. Comparing of Uniq Source IP/day.

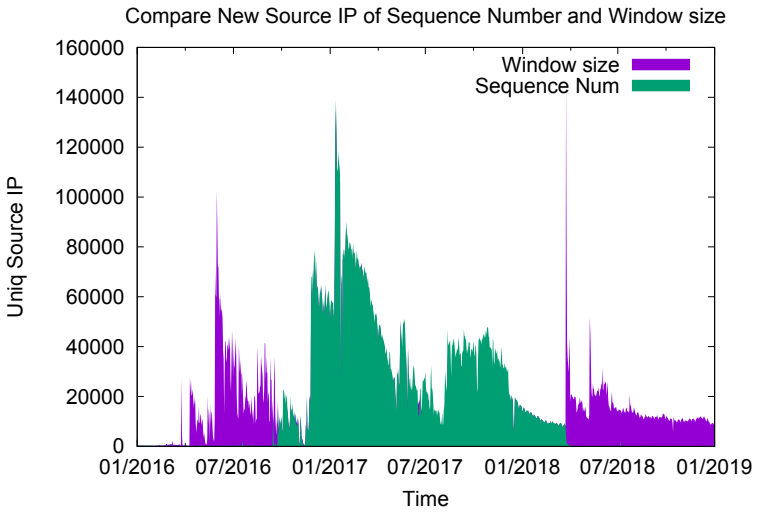

Fig. 9. Comparing of Newly Source IP.

Mirai

Results Based on the Characteristics of Mirai
Mirai was not presented at all until 08/01/2016. Figure 10 shows that it appeared from 08/01/2016 and the infection increased in activity and spread. The peak of the spread of the infection was on 09/21/2016. On that day, about 760,000 hosts participated in the infection activity, and about 380,000 were newly appearing source IP addresses. Since the peak, the infection situation has expanded and shrunk again.
In the second half of 2017 it had shrunk considerably, but on 11/30/2017, we can confirm that there was again a large scale infection activity. On that day, about 680,000 hosts participated in the infection activity, among which about 380,000 were new source IP addresses. The number of new source addresses was found to be almost equal to the number of new senders at the first peak. And after 11/30/2017, infection activity dropped sharply, and Mirai's infections also shrank.

Figure 11 shows that in 2018, tens thousands of new source addresses per day appeared without a sudden spread of infection. In 2018, the largest number of new source addresses per day was about 50,000 on the first day.

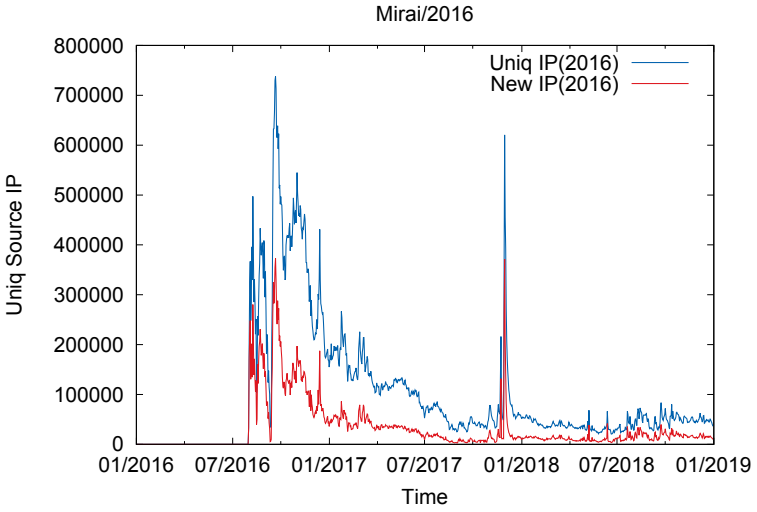

Fig. 10. Spreading of Mirai from 01/01/2016 to 12/31/2018.

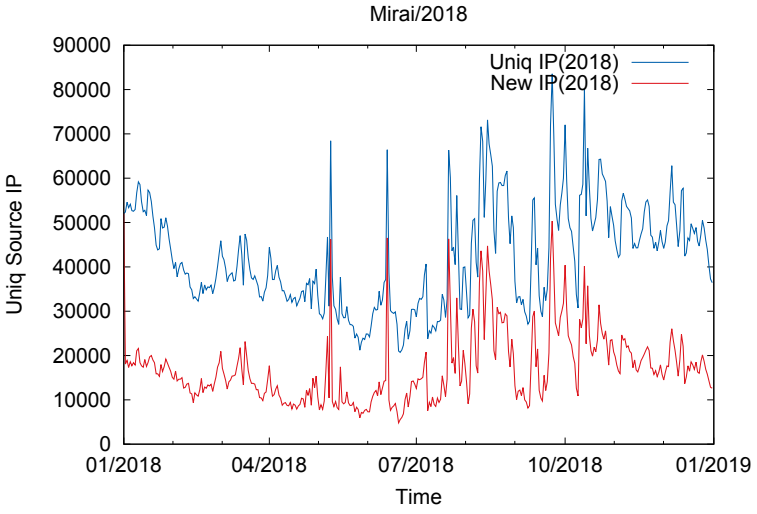

Fig. 11. Spreading of Mirai in 2018.

According to Fig. 12, the trend line for the number of newly appearing source addresses with 2016/01/01 as the observation starting point and the trend line for the number of new source address with 2018/01/01 as the observation starting point appear to almost overlap. New IP(2016) is a transition line for the number of newly appearing source addresses for 2018 with starting point on 01/01/2016. Most of the new source IP addresses in 2018 are real new source addresses and hosts that were infected before 2018 and are not believed to have participated

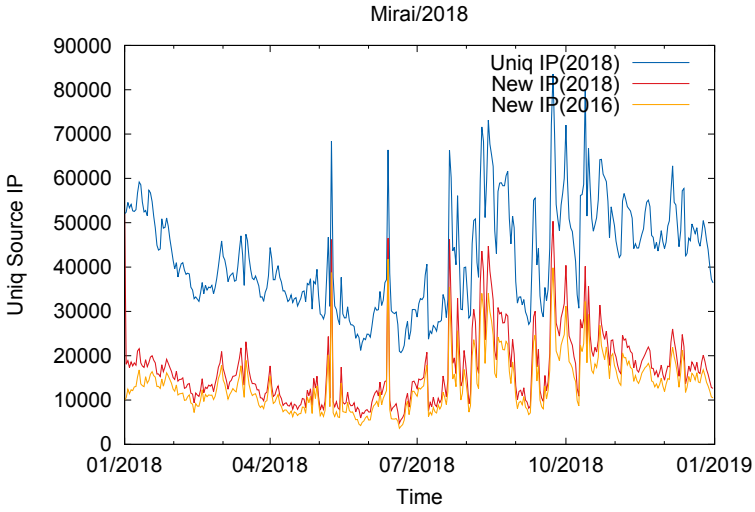

Fig. 12. Spreading of Mirai in 2018 and newly source IP line 2016.

in the infection activity again. Also, the largest number of new source addresses in one day was not 50,000 on the first day, but about 40,000 on May 9, 2018.

5 Discussion

The packets communicated by Hajime were characterized by a fixed window size over a long period of time. Hajime was reportedly infecting telnet with a method of repeated attempts using a prepared list of passwords. Around October 2016, a massive outbreak occurred soon after the appearance of the TCP sequence number feature. At this time, the addition of the TCP sequence number feature and the update of the password list suggested that Hajime had been improved.

As well as April, after the TCP sequence number features were removed and the password list was updated and improved for the second time, large scale infection activity occurred in early April 2018.

We tend to believe that the malware infections will spread along with the increased infection activity of Hajime and Mirai. Looking at the period from Oct. 2017 to April 2018, we can see that while Hajime's infection activity increased, the malware infection shrank. Therefore, an increase in infection activity does not imply an increase in malware infection, it is important to ensure that the number of newly appearing source addresses is investigated to determine the spread and shrinking of the malware.

The reason for Mirai's peak and then decline or increase again is likely due to the release of Mirai's source code, which led to a spike in infection activity each time a variant appeared.

6 Conclusion

This paper analyzes long-term observation data in the darknet. We were able to determine the source addresses of packets with characteristics of Mirai and Hajime, the IoT malware analyzed. We also confirmed that Hajime is a malware that existed long before Mirai, and showed the modified and improved period of the TCP sequence number feature of the packets communicated by Hajime. The results of this study provided a lot of information and clarified the expansion and contraction of malware with characteristic features, thus achieving the purpose of the study.

However, when the experiment was conducted, it was difficult and time consuming to compute due to the very large amount of observed data. Future problems should change the algorithm for this extraction and devise a way to process it efficiently.

References

1. Evans, D.: The Internet of Things - How the Next Evolution of the Internet Is Changing Everything. CISCO White Paper, April 2011. http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf
2. Cui, A., Stolfo, S.J.: A quantitative analysis of the insecurity of embedded network devices: results of a wide-area scan. In: ACSAC 2010, Austin, Texas, 6–10 December (2010). https://doi.org/10.1145/1920261.1920276
3. Costin, A., Zaddach, J., Francillon, A., Balzarotti, D.: A large-scale analysis of the security of embedded firmwares. In: Proceedings of the 23rd USENIX Security Symposium, pp. 95–110, August 2014. ISBN 978-1-931971-15-7
4. Costin, A., Zarras, A., Francillon, A.: Automated dynamic firmware analysis at scale: a case study on embedded web interfaces. In: ASIA CCS 2016, 30 May–03 June 2016. https://doi.org/10.1145/2897845.2897900
5. Crane, C.: Recent Ransomware Attacks: Latest Ransomware Attack News in 2020. https://securityboulevard.com/2020/08/recent-ransomware-attacks-latest-ransomware-attack-news-in-2020/
6. Cyber Security Labo: National Institute of Information and Communications Technology: NICTER Observation report (2017). https://www.nict.go.jp/cyber/report/NICTER_report_2017.pdf, (in Japanese)
7. Threat Hunter Team, Symantec: Petya ransomware outbreak: Here's what you need to know. BROADCOM Symantec Enterprise Blogs/Threat Intelligence. Accessed 24 October 2017. https://symantec-enterprise-blogs.security.com/blogs/threat-intelligence/petya-ransomware-wiper
8. Trend Micro: Bad Rabbit Ransomware Spreads via Network - A ransomware campaign hits Eastern European countries with what seems to be a variant of the Petya ransomware dubbed Bad Rabbit. Trend Micro Research. Accessed 24 October 2017. https://www.trendmicro.com/en_us/research/17/j/bad-rabbit-ransomware-spreads-via-network-hits-ukraine-russia.html
9. Cyber Security Labo, National Institute of Information and Communications Technology: NICTER Observation report 2018. https://www.nict.go.jp/cyber/report/NICTER_report_2018.pdf, (in Japanese)

10. IIJ Group Security Coordination Team: IIJ-SECT Security Report 2018, 28 January 2019. https://sect.iij.ad.jp/d/2019/01/288147.html, (in Japanese)
11. IIJ Group Security Coordination Team: IIJ-SECT Security Report 2019, 4 February 2020. https://sect.iij.ad.jp/d/2020/02/030029.html, (in Japanese)
12. IIJ Group Security Coordination Team: Hajime bot observation status, 1 September 2017. https://sect.iij.ad.jp/d/2017/09/293589.html, (in Japanese)