A Longitudinal Measurement and Analysis of Pink, a Hybrid P2P IoT Botnet

Binglai Wang[1,2], Yafei Sang[1(✉)], Yongzheng Zhang[3], Shuhao Li[1], Ruihai Ge[1], and Yong Ding[1]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
{wangbinglai,sangyafei,lishuhao,geruihai}@iie.ac.cn
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3] China Assets Cybersecurity Technology CO., LTD., Beijing, China
zhangyz@cacts.cn

Abstract. With the ubiquitous deployment of Internet of Things (IoT) devices in many fields, more and more IoT botnets have taken a variety of penetration methods to infect vulnerable IoT devices. Nowadays, a substantial Peer-to-Peer (P2P) IoT botnet named Pink has infected over 1.6 million IoT devices since January 2020, and its impact once exceeded other notorious IoT botnets, such as Mirai, Hajime, Mozi, and so on. Pink is the first IoT botnet using a hybrid topology with centralized and decentralized network architectures. Its two distinct features can be summarized as follows. (i) Different from the conventional P2P IoT botnet based on the public Distributed Hash Table (DHT) service, Pink introduces a novel mechanism called B-segment to build a P2P network, which makes it challenging to track the entire botnet. (ii) Pink is the first IoT botnet to leverage third-party services to propagate configuration files, thereby increasing its resilience. In this paper, we propose an active detection method to measure and understand the development and changes of the Pink botnet continuously. Through daily and continuous measuring of the Pink botnet since January 2022, we firstly provide a comprehensive view of its inapparent network, including bot sizes, global geographic distribution, daily activity, configuration analysis, and Pink botnet countermeasures. We believe that our measurement result is infinitely close to the boundary of the Pink network. Through this study, we reveal that deeper penetration attacks are occurring in the IoT field, and there is an urgent need to improve the security protection of IoT devices. Meanwhile, we hope that this study can promote future research on IoT botnets.

Keywords: Botnet · P2P · C&C · IoT · Pink · Network

1 Introduction

A survey reported by IoT-Analytics reveals that 30.9 billion Internet of Things (IoT) devices are expected to be in extensive use worldwide by 2025 [12,14].

H. Gao et al. (Eds.): CollaborateCom 2022, LNICST 461, pp. 419–436, 2022.
https://doi.org/10.1007/978-3-031-24386-8_23

Unfortunately, these devices with severe flaws are prone to be infected by various IoT botnets, which fundamentally change the internet threat landscape. Although the various attacks and vulnerabilities launched by IoT botnets have been in-depth analyses [4,7,13], there remains much to understand about how the underlying ecosystem work of infected IoT devices. For example: How are the compromised IoT devices geographically distributed? How large can a global IoT botnet grow daily? How does an IoT botnet maintain software updates rapidly and thoroughly? To answer these questions and more, we present an in-depth measurement and analysis of a prevalent IoT botnet, Pink, in this paper.

Pink provides three novel characteristics, which are essential forward-looking significance for studying P2P IoT botnets. (1) Pink is the first IoT botnet with a hybrid topology with centralized and decentralized architectures to distribute attack instructions. The architecture can present high fault tolerance, which can remedy a single point of failure in a central C&C network and distribute configuration files in real-time. (2) Pink is the first P2P IoT botnet to apply an anti-track mechanism. Pink customized P2P communication and mixed it into Network Time Protocol (NTP) service. Then, the messages sent to the other bots are difficult to be detected through network features. (3) Different from Hajime [8] and Mozi [2], Pink is the first P2P IoT botnet built with the B-segment mechanism instead of using the public Distributed Hash Table (DHT) service. The advantage of the B-segment mechanism is two-fold: (i) deeper concealment making up for the defect that the DHT-based mechanism is easy to be tracked; (ii) more flexible network management by probing the alive Pink bot in Class B IPs. We present the details of the B-segment mechanism in Sect. 2. Therefore, the measurement and analysis of the Pink is instructive for understanding and governance of P2P IoT botnets. With this work, our contributions can be summarized as follows:

- **To the best of our knowledge, we are the first to elaborate a comprehensive study of the large-scale Pink IoT botnet.** We investigate the properties of Pink, *e.g.*, network scale, geographical distribution, bot lifetime, communication patterns, and so on.
- **We propose an active measurement method, which can not only track Pink bots but also can be extended to other similar P2P IoT botnets.** The key novelty of this approach lies in two points: (i) It infiltrates the entire Pink botnet by actively sending customized message packets conforming to known Pink bots, and (ii) It passively waits for active interaction from more unknown zombie bots because they can sense the existence of our probe nodes through other Pink bots. Compared with the public-DHT-crawler-based method (adopted to measure Hajime and Mozi), our solution can reduce unnecessary costs and apply to other IoT botnets using P2P DHT communication.
- **We offer two fundamental insights on Pink botnet from our measurement study.** (i) By analyzing the network behavior of prevalent Pink binary samples, we find that P2P communication dominates in distributing config files; we believe that other P2P IoT botnets will adopt the P2P

network construction method based on the B-segment mechanism in the future. (ii) From the perspective of geographical distribution, 99% of Pink bots are located in China and show a B-segment distribution with a long life-time, suggesting that the attack target of the Pink controller is well-defined and widely distributed in China.

2 Preliminaries

Pink is a hybrid architecture botnet with P2P and central C&C communication patterns. This architecture enhances its robustness and facilitates the attacker's control of the entire botnet. In this section, we introduce Pink composition and operation that are most relevant to our study.

2.1 Pink Composition

Through our tracking and analysis, we divided the Pink botnet into five parts: Botmaster, Third-party, P2P, Central Command and Control (Central C&C), and Pink bot. The structure of the Pink botnet is shown in Fig. 1. Different from conventional IoT botnets like Hajime or Mozi that only use the DHT method to complete configuration distribution tasks, Pink botnet has adopted three communication patterns, including third-party, P2P communication, and central C&C, to improve the robustness of the entire botnet. The attacker can operate the three patterns through Botmater to send the configuration files to Pink bots. Each component plays a different role in the botnet. We provide the details in this section.

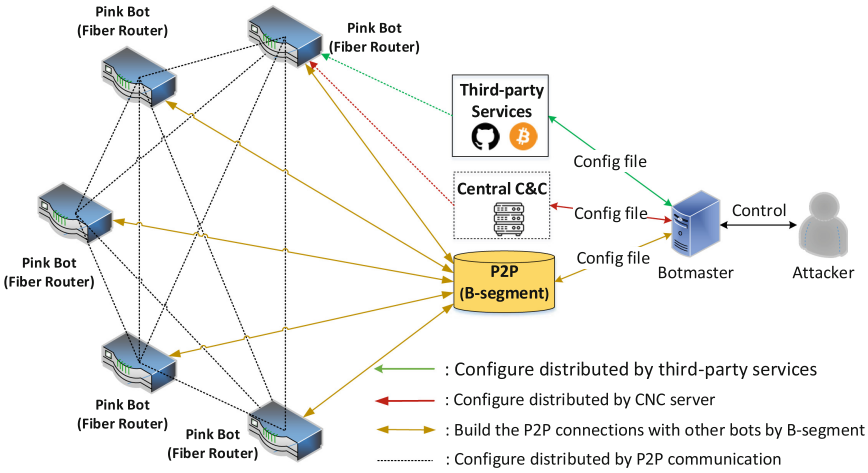

Fig. 1. The structure of Pink botnet.

Botmaster. Pink botnet presents a command relay role, namely botmaster, to send commands to other bots. When the attacker needs a botnet to perform specific actions, it only needs to send a configuration file to the Botmaster. Then, to ensure the correct distribution of the configuration file in the botnet, Botmaster has utilized three communication patterns, including third-party services, P2P, and central C&C, to deliver the configuration file. After receiving the configuration file, the Pink bot will execute the commands defined in the new configuration file and distribute this file to its neighbor Pink bots.

Third-Party Services. Pink botnet leverages another scheme to deliver configuration files. The attacker uses the third-party services Github [10] and BTC [16] to propagate the config file. First, Pink bots can leverage the transaction records in a specific BTC wallet to obtain the topic tags related to the GitHub project. Second, Pink bots will go through the issues of these projects and look for a hidden Git project. This scheme has a robust anti-strike capability. The reason is that the specified BTC wallet must be blocked to disrupt the GitHub-based distribution logic of Pink.

B-segment. Pink botnet provides a unique mechanism that enables new Pink bots to rapidly discover other existing Pink bots and make a connection. Its principle is that the IPv4 address space of many vulnerable IoT devices is distributed in the same Class B network, whose first 16 bits are the network part of the address. Therefore, a considerable number of Pink bots accumulate after infection, resulting in their IP addresses showing the characteristics of specified Class B distribution. Then, with the help of the fixed P2P communication port (Network Time Protocol) in each Pink bot, the new Pink bot can discover other Pink bots by traversing all IP addresses in the several specified Class B networks. Our reverse analysis of Pink binaries reveals that the infected Pink bots tend to firstly launch a peer probe request to four Class B networks, namely 114.25.0.0/16, 36.227.0.0/16, 59.115.0.0/16, and 1.224.0.0/16, with the content '1C 00 00 00' under the intension to make the connections with other Pink bots.

Central Command and Control. Another approach attackers use is to distribute config files through a centralized command-and-control server (cnc.pi--nklander.com) hard-coded in several Pink binaries. Therefore, it is accessible to block the centralized communication scheme by blocking the resolution of the domain name.

Pink Bot. The entire Pink botnet composes of various vulnerable fiber routers based on MIPS architecture. These compromised IoT devices, namely the Pink bot, play the core functional role in the botnet. Its functions can be summarized as follows: (1) As a vital component of the P2P network, it leverages the B-segment mechanism to discover other Pink bots and maintains a neighbor list to disseminate command configuration files issued by the Botmaster hierarchically.

(2) As an executor of a network attack, the Pink bot will follow the instructions to respond to the target with specified attacks like DoS, HTTP message injection, etc. (3) Distinct from the conventional IoT botnet, the Pink bot can flash the original firmware of the fiber router to achieve long-term persistence.

2.2 Pink Operation

Although Pink bot has updated several versions since the first discovery, its overall function and operation mechanism have not been changed significantly. In this section, we conduct a reverse analysis of a sample captured in November 2021 to illustrate the working mechanism of the Pink botnet. We split its behavior chain into three stages: infection, initialization, and maintenance.

Infection. Unlike the propagation method of worm-like IoT botnet Mirai [11], the Pink botnet adopts a novel infection method, namely centralized target scanning. During the infection phase, the Pink botnet controller will look for new targets to infect. First, it will utilize the B-segment mechanism to enumerate all IP addresses and scan the potential new target for specific vulnerabilities. The operation is because many identical IoT devices are distributed in the same Class-B IP and have the same vulnerabilities to exploitation by the attacker. The apparent feature can help the botnet controller quickly discover potential vulnerable IoT devices. Once the potential new target with vulnerabilities is located, the controller will plant the new malicious sample on the new target. In this way, Pink's botnet proliferates and continues to expand. For instance, the attacker leverages the vulnerability originating from misconfiguration in a TCP-17998 control service to gain control of the relevant various fiber routers.

Initialization. The initialization stage aims to join the entire Pink botnet and synchronize the latest configuration file. Pink is an IoT botnet with hybrid network topology, including central C&C and P2P. Therefore, when a Pink sample is planted on a vulnerable IoT device, it will attempt to discover other bots in the P2P network and communicate with the central C&C server in its first execution. Firstly, The Pink bot binds port number 123, commonly used by the NTP service, to communicate with other bots. Subsequently, it can launch various customized probe requests to many IP addresses enumerated from four B-segment addresses ("114.25.0.0/16", "36.227.0.0/16", "59.115.0.–0/16", "1.224.0.0/16") until discovering other active Pink bots and initialize a Pink bot neighborhood table. Simultaneously, through static reverse analysis of Pink samples, we find that the new Pink bot will send a request to the specified central C&C server to acquire the configuration file in the initial stage.

Maintenance. When the attacker needs the botnet to perform specific actions, it only needs to send a customized configuration file with the latest instruction information to any Pink bot through the P2P network or central C&C server.

Table 1. Description of the specified fields in a config file.

Field	Description
verify	Timestamp when the command is issued
cncip/port	Specified IP and port of the latest centralized server
dlc/dl	Pink binary download URL and its Hash check value
sd0/sdp0	Specified DNS server address to resolve a DNS query
srvk	Public key content (base64 encoding) of centralized server
pxy	Specified proxy option

Table 1 presents the required fields in the configuration file. After completing the initialization phase task, the Pink bot needs to maintain communication with other bots to update the neighbor bots table and continuously obtain the new configuration files to take corresponding actions.

2.3 Comparison of IoT Botnets

Table 2. Comparison of IoT botnets

Botnet	Start	Detection[a]	C&C	Protocol	Size	Persistence[b]	Attack[c]
Mirai	2016	E	Centralized	IRC	>100k	W	>1
Bashlite	2014	E	Centralized	IRC	>50k	W	>1
Hajime	2016	M	Decentralized	DHT	>1000k	M	1
Mozi	2019	M	Decentralized	DHT	>1000k	M	2
Pink	2019	C	Hybrid	NTP	>1500k	S	4

[a] Detection indicates the difficulty of distinguish the C&C traffic of different IoT botnet from the abnormal traffic. (E:Easy, M:Moderate, C:Challenging)
[b] Persistence represents the survivability of a Pink bot in a infected IoT device. (W:Weak, M:Medium, S:Strong)
[c] Attacks describe the attack modules in the sample.

In recent years, more and more vulnerable IoT devices have been compromised by various IoT botnets to take network attacks. Table 2 compares Pink and other IoT botnets from seven aspects. Compared with the four widely spread existing IoT botnets, the Pink botnet has a more significant size, more sophisticated C&C channel, more attack modules, more robust scalability, and more complete countermeasures with vendors. The details can be summarized as follows: (1) Although Pink emerged later than other IoT botnets, the number of IoT devices infected by Pink is much higher than other IoT botnets, exceeding 1.5 million. (2) Pink is the first IoT botnet with centralized and decentralized C&C channels. This hybrid network topology can not only make up for the defect of a single point of failure but also have the characteristics of real-time command distribution. (3) From the life cycle perspective, the lifetime of a Pink bot is much longer than that of other IoT botnet nodes; the analysis of Pink binaries reveals that it may be related to the ability that the Pink bot can flash the original firmware of IoT devices and bind UDP 123 port to trick some users into

treating them as a standard NTP service to enhance concealment; (4) From the defense perspective, the Pink bot presents multiple countermeasures to security researchers' mitigation solutions; for instance, the attacker issues the commands to shut down the service through a centralized C&C server, making the vendor unable to patch the compromised IoT devices.

3 Measurement Method

In this section, we propose an active scanning method to infiltrate and measure the entire Pink botnet.

3.1 Active Scanning Method

In bot-scale detection of Pink botnet, we conducted a similar breadth-first search based on the B-segment mechanism, aiming to obtain unknown Pink bots in hierarchical order. Based on the principles and characteristics of the Pink, we figured out that the new Pink bot can join the Pink botnet through the built-in eight startup B-segment addresses, as shown in Table 3.

Table 3. Eight startup B-segment IP addresses hard-coded in the Pink samples

B-segment address	Ports	B-segment address	Ports
114.25.0.0/16	123	61.230.0.0/16	123
36.227.0.0/16	123	110.16.0.0/16	123
59.115.0.0/16	123	118.41.0.0/16	123
1.224.0.0/16	123	211.205.0.0/16	123

The execution flow of our active scanning method is composed of the following three stages. First, we need to initialize a probing table with the above B-segment addresses to enumerate all potential IP addresses to be scanned. After that, our detection bots will construct a customized UDP packet with the payload '1C 00 00 00' and send them to the targets above. Since the sent packets follow the Pink communication pattern, the IP addresses of our detection bots will be added to the Pink bot's P2P communication table and even propagated to other Pink bots. Consequently, there is a certain probability that our probe bots can acquire access packets from other unknown bots in the Pink. To increase the likelihood, we need to send customized packets to as many Pink bots as possible above. At this point, we are unsure whether the access to our detection bots is from the real Pink bots. Then, our detection bots will attempt to parse the received data according to the principle of processing messages by Pink bots. If the target bot returns a config file, it is a real Pink bot, and our detection bots can add the resolved B-segment addresses to the probing table in Step (1) for further processing. If the target bot returns a Network Time Protocol (NTP) packet, it suggests the message is from a standard NTP server.

Algorithm 1. Pink Bot Recognition through Received Messages

Input: *Received_message*: Bytes sent to our measurement program;
Output: *Pink_bot_tag*: The tag represents whether the message's source;*Config_file*: Parse the config file from the received message;

1: **if** $NTP(Received\_message) == True$ **then**
2: // The source sending the messages is a NTP server.
3: Pink_bot_tag=False;
4: return;
5: **end if**
6: **if** $Received\_message.Payload ==' 0000001D'$ **then**
7: // The source sending the messages is a Pink bot without C&C server.
8: Pink_bot_tag=True;
9: return;
10: **end if**
11: **if** $Is\_Pink(Received\_message) == True$ **then**
12: // extract the config file if the received message belongs to Pink
13: Config_file=Pink_decrypt(Received_message);
14: Pink_bot_tag=True;
15: **end if**

We utilize Algorithm 1 to present how to identify the packets related to the Pink bots from the response packets, including NTP and Pink communication packets, respectively. Initially, the first step is to determine the messages of NTP servers through the timestamp feature (line 1). From the reverse analysis of several Pink binary samples, we know that a Pink bot's response message depends on whether it has obtained the C&C server info when receiving a customized UDP packet with content '1C 00 00 00'. If the target bot does not have C&C information, it will respond with '1D 00 00 00'. When the target bot has already gotten the C&C information, it replies with the signature of the C&C data and the corresponding config file. Therefore, we use two branches to deal with the response message from a Pink bot. One is to process the message with payload '1D 00 00 00', which indicates that the Pink sender bot doesn't have any C&C server info (line 6). The other is to extract the config file from the received message and parse the latest instruction info.

3.2 Method Evaluation

According to the key metrics for botnet structures [6], we define two evaluation metrics, namely effectiveness ratio and bots' daily increment, to discuss the rationality of the active scanning method. The former metric is used to verify the effectiveness of the active scanning method, and the latter is applied to present that our approach can adapt to the change in Pink botnet's scale.

Effectiveness. We distinguish the real Pink bot through the feature of whether the communication payload contains a config file. If the received content includes a config file, it can be considered that the bot sending this message is a real Pink bot. We introduce a critical metric *effectiveness ratio*, which depicts the

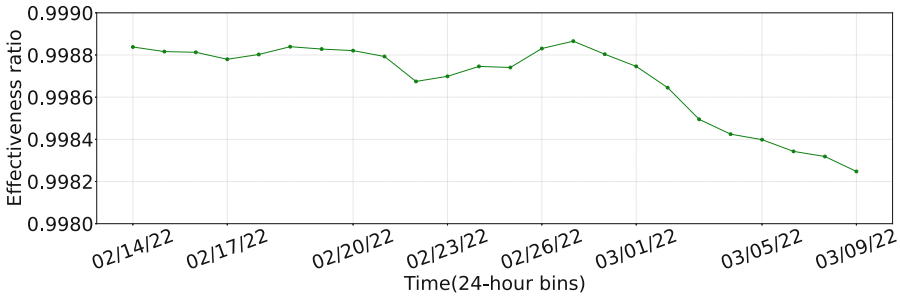

Fig. 2. Effectiveness ratio.

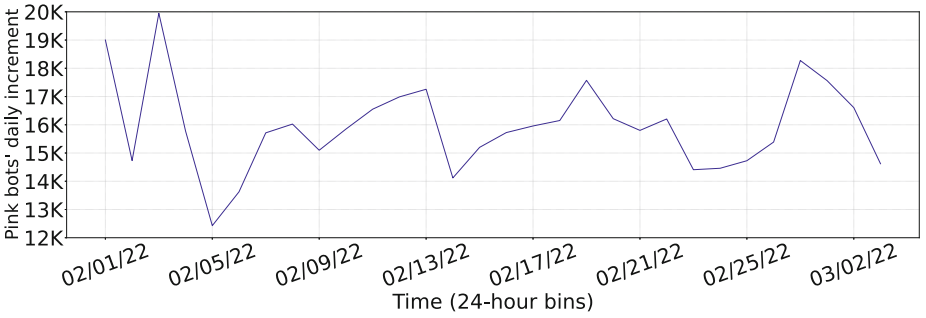

Fig. 3. New Pink bots.

proportion of real captured Pink bots to all collected nodes per day. It is denoted as V_i and can be calculated as follow:

$$V_i = \frac{\Psi(i)}{\Pi(i)} \tag{1}$$

The number i specifies a date, such as 2022-02-14. In the equation, $\Pi(i)$ describes the total number of captured bots perday and $\Psi(i)$ represents the number of real bots filtered per day from $\Pi(i)$. Based on the raw packet data collected by the active scanning method, we calculate the metric V_i from Feb 22, 2022, to Mar 9, 2022, as shown in Fig. 2. Obviously, all days' effectiveness ratio exceeds 0.998, suggesting that our method can collect Pink bots effectively. It should be noted that the tiny gap from 100% represents the proportion of collected standard NTP service, and this phenomenon cannot affect measurement results.

Bot's Daily Increment. Figure 3 shows the number of new Pink bots captured daily by the active scanning method. We cannot guarantee 100% coverage of all Pink bots because some pink bots are on the periphery or even isolated. However, We reckon that the following reasons account for why the detected bots occupy most of the Pink network. First, the number of B-segment addresses to which the Pink bots belong tends to be stable, with almost no new additions. Secondly, the

number of new Pink bots added each day is relatively stable with no significant fluctuation.

The prototype system of our active scanning method is deployed in twenty Virtual Private Servers (VPS), and we report measurements from January 31, 2022, to April 13, 2022. Throughout our active measurement results, we collected 1,542,558 unique IP addresses, most of which are located in regions like China and South Korea. We present a detailed analysis of the Pink botnet's measurement in Sect. 4.

3.3 Ethical Considerations

In the scope of our study, the active scanning method only communicates with the Pink bots' P2P module to obtain the response packets. The operations do not disrupt the bots or the IoT devices on which the bots execute. We have not exploited or infiltrated any compromised IoT devices with misconfigurations or vulnerabilities.

4 Measurement Analysis

We perform a detailed measurement analysis through the acquired bot info about Pink. Since Pink makes full use of the NTP service to build its P2P network, we leverage the detected IP address to identify an infected device.

4.1 One-Day Monitoring

Figure 4 presents the number of online and new bots we capture in each 5-minute interval over a day (02/28/22). The reason why we choose this date is that our scanning system acquires the highest number of online bots and scanning results remain relatively stable all day.

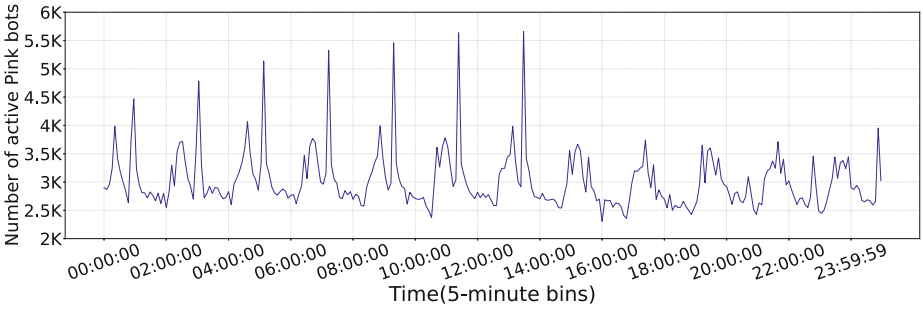

Fig. 4. The number of online Pink bots.

Figure 4 shows that the number of distinct bots presents a periodic cycle state and is maintained at about 2500 to 5000 every 5 min. We attribute this

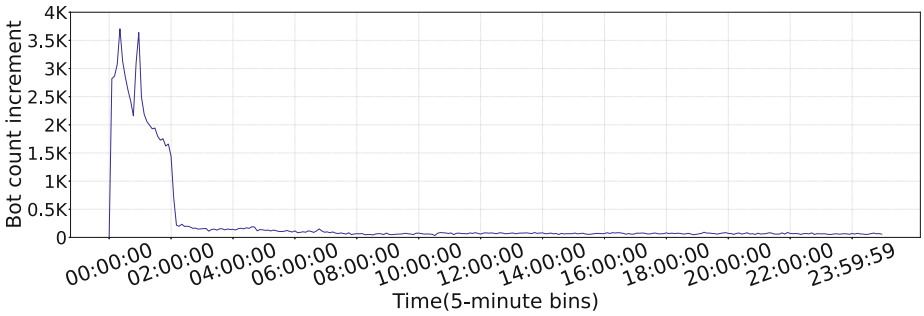

Fig. 5. The number of new Pink bots.

phenomenon to the fact that our prototype system needs to scan the Pink bots table cyclically to determine whether the bot is online. It takes about an hour to complete a full scan. And each small wave peak in Fig. 4 is that our scanning method obtains new reachable bots (unknown Pink bots). Figure 5 presents the change in the number of new Pink bots acquired every 5 min. It reveals that the new Pink bots count during the first period increases rapidly, and when the increment reaches the peak, it drops quickly. Subsequently, the detected increment has gradually stabilized between 50 and 100. If we regard Pink networks as a graph, the area we acquire within two hours may only be the densest and the most accessible part. As for the Pink bots on the periphery or temporarily isolated, we need to continuously scan the obtained Pink bots to collect the message from the remaining unknown Pink bots as much as possible. The smooth increment in the latter phase of Fig. 5 reveals the number of hard-to-scan bots in the Pink.

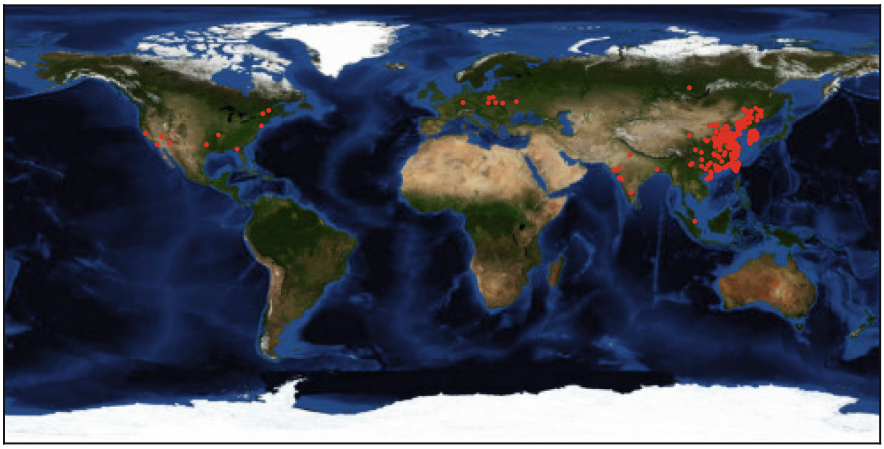

Fig. 6. The geographical distribution of infected devices.

4.2 Continuous Monitoring

Under the premise of th approach in Sect. 3.1, we have taken statistics on the total number of Pink bots from January to April 2022. To understand where Pink infections were geographically concentrated, we extracted the recorded IP addresses from the collected data and calculated the distribution of Pink bots by mapping these addresses to geographic locations. Figure 6 presents the total number of compromised devices in the primary two countries from January to April 2022. Obviously, the geographical distribution that the bulk of Pink infections stemmed from devices located in China (99.10%) and South Korea (0.89%), and the total number of the infected devices accounts for nearly 99.99% in the above two countries. The remaining infection devices in the other countries do not even exceed 0.01%, far less than the number of compromised devices in China and South Korea. Compared with most bots infected by Mozi botnet are diverse and widely distributed worldwide, the primary target devices of Pink infection are fiber routers distributed in China. It suggests that the author of Pink botnet was concerned about potential avenues in China when it was designed.

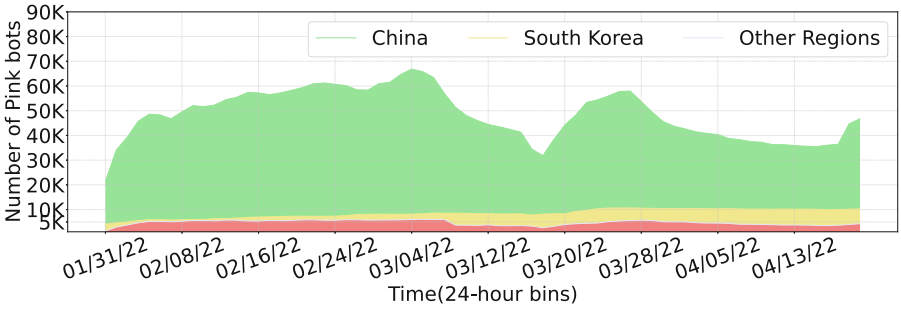

Fig. 7. The daily active bots in different regions.

Figure 7 shows the number of active Pink bots in different regions. The number of active Pink bots collected per day in China is around stable 30K and 70K. It is worth noting that Fig. 7 shows a slow recovery after a clear downward trend in the number of active Pink bots around mid-March 2022. We combine Fig. 5 to track Pink and attribute this phenomenon to the fierce competition between attackers and security researchers for compromised devices. The vulnerability under attack originated from a TCP-17998 control service, an interface for vendors to operate the machines. Since misconfiguration of the service leads to open access in the public network, the attacker gains control of the relevant fiber routers. Then, device vendors, with the assistance of a cybersecurity company, attempt to fix the compromised devices through the above service [3]. However, the attacker sends a message to close the TCP-17998 control service through the propagated config file in these compromised devices, cutting off the vendor's control over the devices. Finally, the only option left for the vendor is to physically access the fiber router, disassemble the debugging interface or replace

the unit. And the number of active Pink bots detected in our two figures reflects the intensity of the war during the period.

4.3 Bot Analysis

Birth and Death. We conduct an in-depth tracking and statistical analysis of Pink in Sect. 4.1 and Sect. 4.2. Figure 8 presents the number of births and deaths about Pink bots per day from 3 February 2022 to 2 March 2022. The births imply the new Pink bots are obtained every day, and deaths mean that known Pink bots are not active now (the infected devices have been repaired by vendors or have been offline due to network reasons). We can find that the number of Pink bots generated accounts for a quarter of the total detected online bots per day in the botnet, with the number of extinguished Pink bots gradually increasing. It suggests that the entire Pink botnet is in a significant dynamic change. During the measurement period, the number of Pink's births just started to present a downward trend and remained stable, while the number of deaths offered a growth trend. The reason is that vendors and the cybersecurity community have been working on methods to govern the Pink botnet. With the efforts of various vendors, we firmly believe that the number of Pink's deaths will increase significantly, and Pinks' births will gradually decrease.

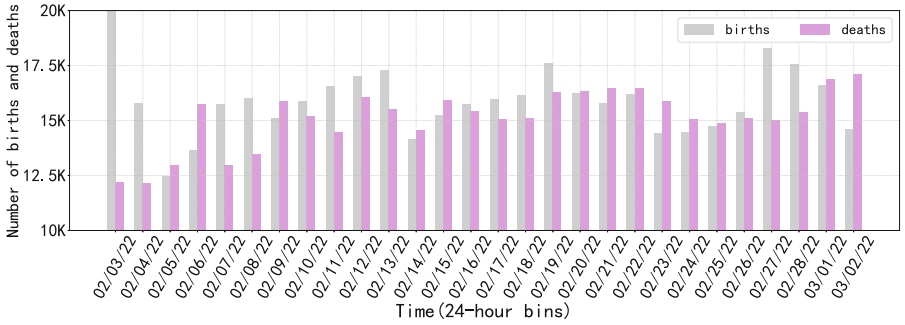

Fig. 8. The number of birth and death bots in Pink botnet for a day.

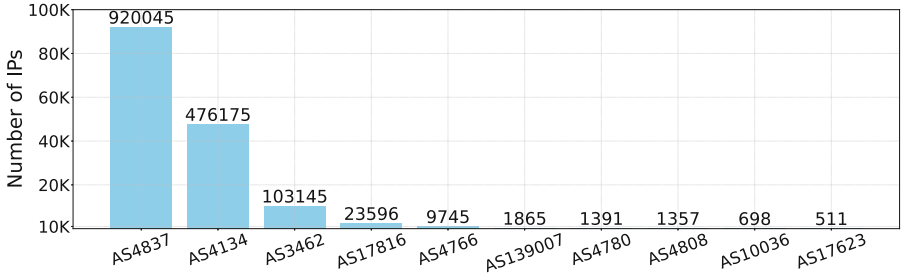

Fig. 9. Top ten ASes ranked based on the number of Pink bot infected IP addresses.

Bot ASes Distribution. These observations stemmed from 1,542,558 distinct IP addresses located across 15 Autonomous Systems (ASes) spanning several regions between January 2022 and April 2022. As already mentioned, we utilize MaxMind's[1] GeoLite database to explore the geographic distribution of collected Pink IP addresses and find that 99% of them are located in China. It suggests that China is the region most affected by this botnet. During the initial execution of a new Pink bot, it constantly scans various IP addresses from B-segment. It then attempts to establish connections with other compromised bots in the Pink P2P botnet. As shown in Fig. 9, we map millions of compromised devices' IPs to their corresponding Autonomous Systems (ASes). It reveals that 99% of the infected IPs in our collected data only reside in the top three ASes, mainly in China. Evidently, the remaining 1% of Pink bots are spread sporadically over random Ases spanning 15 regions. In conclusion, we can conclude the two interesting findings about the Pink botnet. (1) The whole Pink botnet is primarily composed of a single type of fiber router located in China, suggesting that the attacker has a perfect understanding of the exploit and distribution of these embedded devices. (2) The IP addresses of the compromised devices exhibit an aggregated B-segment distribution feature, allowing attackers to build an advanced P2P IoT botnet through enumerating and probing various IP addresses in the B-segment table.

4.4 Evaluation

Since September 2019, Pink has undergone several iterative updates and has gradually become a million-scale IoT botnet. To understand the possible network behavior and operations of malicious Pink samples, we build an IoT sandbox virtual environment to execute Pink binary samples and analyze the dynamic behavior through the network traffic [15]. Through monitoring and analyzing Pink's binary samples, we hope to answer the following two questions: (1) How does Pink maintain the continuous operation of the entire botnet through network behavior? (2) During the measurement period, how often is the configuration file updated, and what can we learn from the frequently changed config file?

Table 4. Several key fields in the common captured config file (URL1–https://gitee.com/ghy8/bh/raw/master/dlist.txt)

verify	cncip1	cncport1	dl	sd0	sdp0	pxy
1646064005	78.141.194.8	35662	http://217.69.5.95:8010/dlist.txt	URL1	443	1
1640971039	140.82.40.29	26022	http://209.250.247.60/dlist.txt	URL1	443	1
1646064001	78.141.194.8	26022	http://80.240.25.98/dlist.txt	URL1	443	1
1640971040	140.82.40.29	26022	http://217.69.5.95/dlist.txt	URL1	443	1
1648784552	78.141.194.8	35778	http://217.69.5.95:8010/dlist.txt	URL1	443	1

[1] MaxMind: http://www.maxmind.com/en/home.

Communication Traffic and Network Behavior. We leverage IoTPoT to execute several prevalent Pink binaries and collect these samples' communication data [15]. As described in Sect. 2.1, Pink employs three communication modes to propagate the config file in the botnet. Table 5 presents the proportion of communication protocol packets in a Pink sample. From the Table, we can learn that more than 85% of the communication behavior belongs to UDP protocol, and only less than 15% is TCP protocol. The reason is that the Pink botnet primarily constructs a P2P botnet through a customized UDP protocol (78.96%) to maintain communication among millions of bots. From the traffic analysis results, we figured out that Pink bots were attempting to send customized UDP packets to many enumerated IPs from B-segment addresses and waiting for the responses. Since there must be a small number of Pink bots addresses in the B-segment, traversing the IPs makes it bound to find several of them. Then the bot can discover other active Pink bots and join the entire botnet. TCP-based P2P and C&C communication are the other primary methods of acquiring config files. It is worth noting that the TCP connections (5.69%) among Pink bots depend on whether the P2P connections based on UDP protocol have been established, indicating that UDP-based P2P connections dominate the entire botnet communication. As for the centralized C&C communication, we can find that the server address that the Pink sample communicated with has become invalid, suggesting that updating config files through a centralized server is unreliable. In summary, the above three network behavior of Pink bots is to acquire config files from other Pink bots or centralized server.

Table 5. Distribution of Pink Communication Protocol. (UDP, User Datagram Protocol; TCP, Transmission Control Protocol; C&C, Command and Control Private Protocol;)

Protocol	Packets	Proportion
P2P-UDP	43141	*78.96%*
P2P-TCP	3109	*5.69%*
DNS	3494	*6.4%*
C&C	4891	*8.95%*

Configuration Analysis. The config file dominates the update of binary samples and the delivery of commands in the entire Pink botnet. Therefore, the analysis of config file updates is critical to understanding the function of the entire Pink botnet. As described in Sect. 2.1, we know that several fields (cncip1, cncporta, dl, and sd0) are required during the propagation of the configuration file in the Pink botnet. Table 4 presents several key fields in the prevalent config file collected during the measurement period. The dl field provided a specified download URL of the sample update, and we only captured three changes during the measurement period. It is worth noting that the download address of this field is prone to invalidation. We can speculate that the main reason is that the

download server may have been countered by security researchers or shut down by attackers.

5 Related Work

Over the last decade, attacks from IoT botnets and their variants gradually became the primary threat to IoT devices [18]. This phenomenon attracted lots of attention from the security community, focusing on the measurement, analysis, mitigation, and disruption of IoT botnets [20]. In 2017, Antonakakis et al. leveraged a diverse set of vantage points, including network telescope probes, Internet-wide banner scans, IoT honeypots, command and control (C&C) milkers, DNS traces, and logs provided by attack victims, to conduct a broad study of the Mirai botnet [4]. It was the first formal step to studying and understanding IoT botnet comprehensively. However, these measurement methods had limited effect on understanding and analyzing decentralized P2P IoT botnets.

Other closely related IoT botnets to Pink were the widely studied Hajime [7, 9] and Mozi [1,2,19]. Both used an existing Kademlia-based DHT to distribute C&C information, and similar active DHT measurements were performed to track the above P2P IoT botnets. According to Hajime's DHT design, Herwig et al. provided over a year of retrospective measurement analysis of Hajime, including its size, its C&C infrastructure, its evolution, and the compromised IoT devices [8]. Tengfei et al. also measured the quick spread of the Mozi botnet through a similar breadth-first search based on the topological structure of the DHT network [19]. Obviously, the detection method of the two studies above was effective for measuring P2P IoT botnets based on the DHT OVERNET network but was not suitable for the non-DHT P2P IoT botnet like Pink; we extend it in several key ways.

First, Pink represents a step in the evolution of the P2P IoT botnet in that it leverages the B-segment mechanism (described in Sect. 2.1) instead of the traditional DHT method to build its sophisticated P2P C&C infrastructure. To obtain the infrastructure, we leverage Pink's P2P design to infiltrate the entire botnet and attract messages from other unknown Pink bots. Second, Pink often updates the payloads and incorporates new attack vectors. Using the collected data, we analyze Mozi bots' geographical dispersion and explore the impact of payload updates on the botnet size, location, and composition. Finally, intending to mitigate the attacks from Pink, we summarize its network and distribution features to speculate the possible reasons why the Pink botnet has been active since 2020.

The most immediately related prior works were the studies of the Pink botnet performed by 360 Netlab [3], NSFOCUS [17], and Cyware [5] in the wake of the Pink discovery. The previous studies primarily involved short-term P2P network measurements and reverse engineering of botnet payloads. By comparison, we achieve more prolonged and more comprehensive studies of Pink, allowing us to observe the distribution of bots, the lifetime of various bots, and the impact of payload updates on the botnet.

6 Conclusion

The Pink botnet was first discovered and analyzed by 360 Netlab in January 2020 and has been developed for nearly two years [3]. During this period, we have witnessed its peak development and compromised over 1.6 million devices, most of which are located in China. Pink adopts a novel P2P network establishment mechanism, which needs to brute force the IP addresses in several specified Class B IP addresses. Compared with the previous mechanism based on the public DHT service, it can accelerate the establishment of a P2P network and make it challenging to track the entire Pink botnet, enhancing the hidden ability. Throughout the in-depth analysis of the communication protocol of Pink bots, we propose an active scanning method to simulate some nodes that can communicate with Pink bots to attract responses from other unknown Pink bots. By continuously monitoring the online bots and obtaining more unknown bots, we conduct a comprehensive analysis of Pink's emergency and growth, geographical distribution, the composition of communication data, and the commands in the configuration file. We hope these findings can serve as an alarm for vendors and security researchers to improve the patching of vulnerable IoT devices.

Acknowledgment. We thank the anonymous reviewers for their insightful comments. This work is supported by The National Key Research and Development Program of China (No. 2019YFB1005201, No. 2019YFB1005203 and No. 2019YFB1005205).

References

1. Alex, T., Hui, W., Genshen, Y.: Mozi is dead and the poison remains (2021). https://blog.netlab.360.com/the_death_of_mozi_cn/
2. Turing, A., Wang, H.: Mozi, another botnet using DHT (2019). https://blog.netlab.360.com/mozi-another-botnet-using-dht/
3. Turing, A., Wang, H.: Pink, a botnet that competed with the vendor to control the massive infected devices (2021). https://blog.netlab.360.com/pink-en/
4. Antonakakis, M., et al.: Understanding the MIRAI botnet. In: 26th USENIX security symposium (USENIX Security 2017) (2017)
5. Cyware: Experts disclose pink botnet amidst multiple DDoS alerts (2021). https://cyware.com/news/experts-disclose-pink-botnet-amidst-multiple-ddos-alerts-662ed0c4
6. Dagon, D., Gu, G., Lee, C.P., Lee, W.: A taxonomy of botnet structures. In: Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007), pp. 325–339. IEEE (2007)
7. Edwards, S., Profetis, I.: Hajime: analysis of a decentralized internet worm for IoT devices. In: Rapidity Networks, Security Research Group, Technical report (2016)
8. Herwig, S., Harvey, K., Hughey, G., Roberts, R., Levin, D.: Measurement and analysis of Hajime, a peer-to-peer IoT botnet. In: Network and Distributed System Security (NDSS) Symposium (2019)
9. Van Der wiel, J., Vicente Diaz, Y.N.: Hajime, the mysterious evolving botnet (2017). https://securelist.com/hajime-the-mysterious-evolving-botnet/78160/

10. Kalliamvakou, E., Gousios, G., Blincoe, K., Singer, L., German, D.M., Damian, D.: The promises and perils of mining github. In: Proceedings of the 11th Working Conference on Mining Software Repositories, pp. 92–101 (2014)
11. Kambourakis, G., Kolias, C., Stavrou, A.: The MIRAI botnet and the IoT zombie armies. In: IEEE Military Communications Conference (MILCOM) (2017)
12. Lueth, K.L.: State of the IoT 2020: 12 billion IoT connections, surpassing non-IoT for the first time (2020). https://iot-analytics.com/state-of-the-iot-2020-12-billion-iot-connections-surpassing-non-iot-for-the-first-time/
13. Marzano, A., et al.: The evolution of bashlite and Mirai IoT botnets. In: 2018 IEEE Symposium on Computers and Communications (ISCC), pp. 00813–00818. IEEE (2018)
14. Meulen, R.v.d.: Gartner says 8.4 billion connected "things" will be in use in 2017 up 31 percent from 2016. In: Gartner. Letzte Aktualisierung (2017)
15. Pa, Y.M.P., Suzuki, S., Yoshioka, K., Matsumoto, T., Kasama, T., Rossow, C.: IoTpot: a novel honeypot for revealing current IoT threats. J. Inf. Process. **24**(3), 522–533 (2016)
16. Sidhu, J.: SysCoin: a peer-to-peer electronic cash system with blockchain-based services for e-business. In: 2017 26th International Conference on Computer Communication and Networks (ICCCN), pp. 1–6. IEEE (2017)
17. Team, C.: Experts disclose pink botnet amidst multiple DDoS alerts (2021). https://cyberintelmag.com/malware-viruses/pink-botnet-malware-infected-more-than-1-6-million-devices-according-to-researchers/
18. Trendmicro: IoT botnet (2016). https://www.trendmicro.com/vinfo/us/security/definition/iot-botnet
19. Tu, T.F., Qin, J.W., Zhang, H., Chen, M., Xu, T., Huang, Y.: A comprehensive study of mozi botnet. Int. J. Intell. Syst. (2022)
20. Vu, S.N.T., Stege, M., El-Habr, P.I., Bang, J., Dragoni, N.: A survey on botnets: incentives, evolution, detection and current trends. Future Internet (2021)