Characterisation of the Kelihos.B Botnet

Max Kerkers, José Jair Santanna, Anna Sperotto

Characterisation of the Kelihos.B Botnet

Max Kerkers, José Jair Santanna and Anna Sperotto

Design and Analysis of Communication Systems (DACS)
University of Twente
Enschede, The Netherlands
m.kerkers@student.utwente.nl, j.j.santanna@utwente.nl,
a.sperotto@utwente.nl

Abstract. Botnets are organized networks of infected computers that are used for malicious purposes. An example is Kelihos.B, a botnet of the Kelihos family used primarily for mining bitcoins, sending spam and stealing bitcoin wallets. A large part of the Kelihos.B botnet was sinkholed in early 2012 and since then bots are sending requests to controlled servers. In this paper, we analyze and characterize the behavior of Kelihos.B. Our analysis is based on the log file of the bot request logged at the sinkhole from March 2012 to early November 2013. We investigate both the overall characteristics of the botnets, as well as on its evolution over time since the time of the sinkholing. Our results indicate that, although this trend is decreasing, there are possibly still newly infected bots even more than a year from the original sinkholing.

Keywords: Botnet, Kelihos.B, Hlux2, Characterisation, Sinkhole

1 Introduction

Botnets are one of the modern threats to society. A botnet consists of several malware-infected computers (bots) that are controlled by the owners of the botnet. Botnets have for example been used to send spam or to launch distributed denial of service (DDoS) attacks [1]. One of such botnets is Kelihos.B, which was primarily used for mining bitcoins, sending spam and stealing bitcoin wallets [2].

The Kelihos.B botnet was sinkholed on the 21st of March 2012 by security experts from Kaspersky, CrowdStrike, and SURFnet, among others [2]. Sinkholing is the operation of re-directing C&C requests to a set of controlled servers by reverse-engineering the botnet C&C mechanisms. As these controlled servers will not send any jobs to the bots, the botnet is in practice disrupted [4]. Since

It should be noted that an initial version of this paper has been presented at the 20th Twente Student Conference on Information Technology, as for requirement of the Bachelor degree in Computer Science. However, the conference was an internal event of the University of Twente, of which the proceedings have not officially been published by any publisher [3].

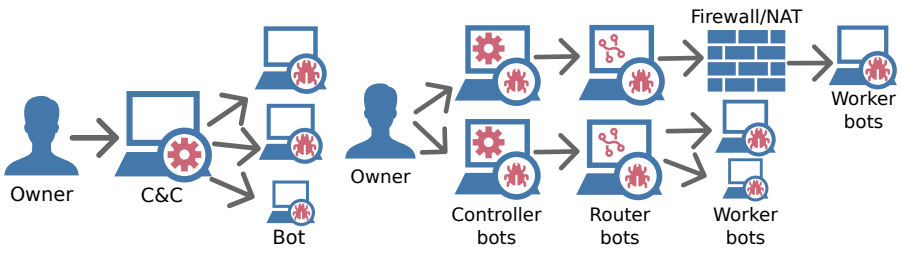
then, the sinkholing servers have collected large numbers of data from the bots that are still active, although ineffective.

The goal of this paper is to characterize the Kelihos.B botnet. This characterization will contribute to the better understanding of the evolution over a period of one year and nine months of a sinkholed botnet (from March 2012 to November 2013). Our analysis confirms and updates preliminary results on the Kelihos.B botnet that have been carried on days after the sinkholing [2, 4] or in occasion of other sinkholing operations [5]. However, the analysis presented in this paper vastly extends previous results and sheds new light on aspects that are not considered in literature. First, we analyze the long term behavior of the hosts contacting the sinkhole, which is only hinted at [5]. Second, we identify the presence, still today, of newly infected hosts from different the Autonomous Systems, which indicates that the infection vector for this Botnet is still partially active. Last, we analyze the temporal behavior of hosts in different continents, which highlights the presence of a clear day-night pattern and implies a certain type of infected hosts.

This paper is structured as follow. In the Section 2 we provide background information about the structure of the botnet and the sinkholing operation. Then the considered dataset and analysis methodology are described in Section 3. In Section 4, the analysis results are provided. Finally, we conclude in Section 5.

2 Background

A botnet is a collection of hosts, called bots, which are infected with malicious software. Bots are typically controlled by one or more Command & Control servers (C&Cs), which belongs to the owner of the botnet and are used to send bots tasks they have to execute [1].



(a) Classical botnet (b) P2P botnet (like Kelihos.B)

Fig. 1: Botnet architectures

In a classical botnet, as shown in Figure 1a, the C&C server is centralized and it has knowledge about the addresses of every bot in the botnet. Despite the simplicity of structuring these botnets, a central C&C server is also a single

point of failure, since only the central server has to be disabled to shutdown the botnet. Recently, botnets based on a peer-to-peer (P2P) infrastructure have become more common. Examples are the Zeus (P2P variant) [6], Sality [7] and Kelihos.B, the botnet considered in this paper. An exemplified architecture of a P2P botnets is shown in Figure 1b. In these botnets, the bots themselves will propagate C&C instructions, and the presence of a centralized C&C server is therefore not needed anymore. When a host is infected and becomes a bot, it will contact a hit-list of trusted hosts from which to request further information and instructions. This process is called bootstrapping.

Kelihos.B is not the first version of the Kelihos botnet to be sinkholed. The first Kelihos botnet (or Hlux) was shutdown by a sinkhole operation in September 2011 [8]. The Kelihos family consisted of three types of bots: controllers, routers and workers, indicated in Figure 1b. The controllers are the bots that are operated by the owners of the botnet and where the instructions originate. The router bots have a list of other bots in the network and their goal is to redistribute the instructions they receive from the controllers. Finally the workers are the bots that execute these instructions. Bots that are located behind firewalls or NATs and therefore cannot be reached from outside are always put themselves in worker mode. Examples of possible instructions these workers can execute include sending spam, participating in distributed denial of service attacks and updating itself to a newer version. These instructions are called jobs. When a bot wants to obtain new instructions it sends a job request to a set of predefined routers. When the bot wants to obtain an update of the list of other bots in the network, it sends a bootstrap request to exchange their lists of routers. Bots that are new in the network always contain a list of several router bots to obtain more router addresses from.

Kelihos.B, a new slightly altered version of Kelihos, was discovered merely weeks after the first sinkholing. Kelihos.B had mostly the same functionality as the original Kelihos botnet. However the botnets also showed some differences, such as: the use of encrypted communication protocol and a new set of encrypted keys. In addition, according [9] and [2], Kelihos.B was used for different purposes, such as to intercept passwords, stealing bitcoin wallets, sending spam and performing DDoS attacks. The family of Kelihos botnets targets Windows-based hosts and it mainly spread via social networks, in particular via a so-called Facebook worm that allures users to download a photo album [10]. At the time of writing this paper it is known that Kelihos has mutated again into a new version that shows the additional feature of stealing Internet browsers passwords [11, 12] and was sinkholed in 2013.

In Section 3 the dataset of the sinkholed Kelihos.B botnet is described.

3 Dataset and Analysis Methodology

The dataset consists of log files from the controlled sinkholing server from the National Research and Education Network of the Netherlands, SURFnet. The log files contain a record for each request to the server done by a bot. Figure 2 is

an example of what those records look like. As can be seen each request is logged together with a human readable timestamp of when the request was received, followed by whether it was a bootstrap request or a job request, from which IP-address and port the request originated. Furthermore, job requests also contain the version of the bot from which the request originates and the operating system of the infected host.

```
[2012-03-21 17:40:27.48661] bootstrap request from x.x.x.x:3810
[2012-03-21 17:40:27.58262] job request from y.y.y.y:2924 - 376ae8[...],
       v126 "plus001", os info: 5.1.2600, platform 2
```

Fig. 2: Examples of log entries (IP-addresses are anonymized)

The dataset spans over a period of one year and nine months, from March 21, 2012 until November 7, 2013. In total, the dataset contains almost 594 million of requests. Therefore this data first had to be aggregated and structured such that it would be easier to analyze.

The logs have been augmented with additional information, such as geolocation data as the originating country and continent of the request, and routing data, i.e., the originating autonomous system. For determining the country from which the IP-addresses originate the MaxMind [13] database was used. In this database the country in which IP-addresses are located can be found. Then, for determining in what continent the country was located the incf.countryutils [14] module was used. This contains a list of which countries are in which continents. For determining the autonomous systems the PyASN module [15] was used. This module uses BGP RIB data from the first of January 2013 to determine which autonomous system an IP-address is in.

The augmented dataset allows us to investigate the following characteristics:

- Types of requests (bootstrap or job);
- IP-addresses and the information derived from those, such as Autonomous System numbers and Geo-location information;
- Port numbers;
- Operating systems (for job requests);
- Bot versions (for job requests);

Our analysis methodology is structured along two main aspects. First, we analyze the dataset as a whole, therefore presenting an *overall analysis* of the main data characteristics (Sec. 4.1). Second, we show a *temporal analysis* of the data, in which we highlight how the sinkholed botnet is evolving over time (Sec. 4.2).

4 Analysis Results

This section presents the outcome of our analysis. First, we present the overall characterization of the Keilhos.B network in Section 4.1; then, in Section 4.2,

we proceed to describe how the botnet has evolved over time in the considered period.

4.1 Overall analysis

The dataset shows that in total 3.7M unique IP-addresses have contacted the sinkhole with 593.4 M requests. Of those, 81.5% were bootstrap requests and 18.5% were job requests. This seems to indicate that the maintenance of the botnet infrastructure, achieved by acquiring bootstrap information, is prevalent comparing with requests for new jobs. This behavior is also ensuring that the botnet, once sinkholed, is continuously controlled, even if new hosts will be infected and became part of the botnet. However, we are not able to quantify if the sinkholing operation has in a way altered the bot behavior, for example causing a larger number of bootstrap requests.

The analysis of the geographical characteristic of the hosts contacting the sinkhole shows that a large fraction of the infected population was located in Poland. This holds both if we consider the percentage of requests received by the sinkhole (Table 1a), as well as the percentage of involved IP addresses (Table 1b). However, Table 1a and Table 1b also indicate that, although several countries appear in the top 10 countries in both the percentage of sent request as well as the percentage of involved IP addresses, there is not a direct correlation between the two lists. Examples are given by the United States, which generate 6.97% of the requests (41.3 M requests), but only hosts 2.24% of the IP addresses (83K IP addresses); or Hungary, which generated 3.66% of the requests (21M requests), but does not appear in the top 10 most active countries in term of IP addresses.

Country	**Request**[%]
Poland	34.13%
United States	6.97%
Turkey	6.85%
Hungary	3.66%
Mexico	3.38%
Argentina	3.24%
Spain	3.06%
Romania	3.04%
Bulgaria	2.22%
Vietnam	2.02%
Others	31.13%

(a) Requests per country

Country	**IP**[%]
Poland	24.85%
Turkey	11.08%
Thailand	4.76%
India	4.74%
Mexico	4.50%
Egypt	4.04%
Vietnam	3.98%
Pakistan	3.56%
Argentina	2.48%
United States	2.24%
Others	33.75%

(b) IPs per country

Table 1: Geographical distribution of request and IP addresses

To better understand how many requests each IP address performs towards the sinkhole, Figure 3 shows the cumulative distribution function of the number of requests per IP address. From the figure, we derive that 17% of the IP

AS	Name	Requests (%)
5617	TPNET Telekomunikacja Polska S.A.	7.28%
9121	TTNET Turk Telekomunikasyon Anonim Sirketi	5.09%
6830	LGI-UPC Liberty Global Operations B.V.	3.78%
29314	VECTRANET-AS VECTRA S.A.	3.23%
21021	MULTIMEDIA-AS Multimedia Polska S.A.	3.20%
12741	INTERNETIA-AS Netia SA	2.50%
7922	COMCAST-7922 - Comcast Cable Communications, Inc.	2.00%
8151	Uninet S.A. de C.V.	1.91%
8048	CANTV Servicios, Venezuela	1.61%
10481	Prima S.A.	1.36%
Others		68.03%

Table 2: Top 10 of Autonomous Systems in percentage of requests

addresses (0.63M IP addresses) have only sent a single request to the server. In addition, 90% of the IP addresses (3.3M) appear up to 135 times in the logs. The remaining 10% of the distribution shows a long tail, where a handful of IP addresses are responsible for up to 2M requests. The most active IP address is located in Montenegro and it created 0.35% of all the requests in the dataset (2M requests).

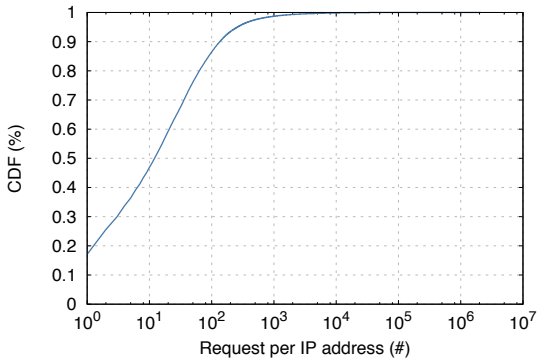


Fig. 3: Cumulative distribution function of the number of requests per IP-addresses

The total number of unique Autonomous Systems (AS) in the dataset is 7629. Also considering ASes, it is easy to see that Poland dominates the top 10 list of most active ASes in terms of requests. As Table 2 shows, the AS where most requests originate from is AS5617 (TPNET Telekomunikacja Polska S.A.), with a share of 7.3% requests (43.3M requests). With a share of 5.1% (30.2 M requests), the second largest origin of requests is the AS9121 from Turkey.

Version	Requests [%]
121	1.47
122	1.12
125	22.65
126	72.93
Other	1.83

(a) Bot version

Operating System	Requests [%]
Windows XP + Windows Server 2003	87,73%
Windows Vista + Windows Server 2008	2,10%
Windows 7 + Windows Server 2008 R2	10,17%

(b) Operating Systems

Table 3: Bot information

Figure 4 shows the cumulative distribution function of the source ports from which the requests originate. The ports that are by far used most are in the range 1024 – 5000, used in 70% of the requests. In only less than 1% of the requests, the bots used to communicate a port in the restricted range (smaller than 1024). The ports in the ranges 5000 – 50000 and larger than 50000 are both equally used in are used in 15% of the requests.

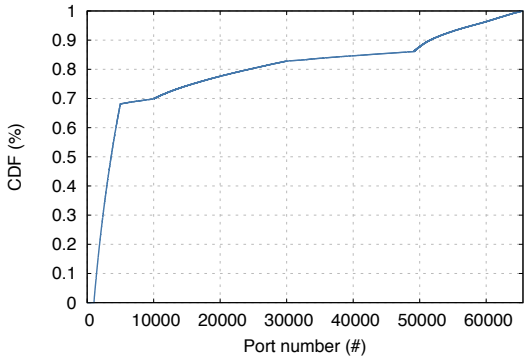


Fig. 4: Cumulative distribution function of originating ports of requests

The two most occurring bot versions are the version 126 with 72.9% and the version 125 with 22.7%. All the other versions combined have a share of less than 5%. This can be seen in Table 3a. As will be seen in the next section these percentages remain close to this values over time.

Finally, our analysis indicates that all requests originate from computers running the Windows operating system. As can be seen in Table 3b almost 88% of the job requests originate from computers running Windows XP (or Windows Server 2003), while the other infected hosts run other operating systems in the Windows family.

4.2 Temporal analysis

In Figure 5 we show the time series of the number of requests per hour received by the sinkhole. The figure shows ranges where data are missing, due to some corrupted logs in the dataset. The overall trend indicated by Figure 5 is an almost exponential decrease in the number of requests per hour, with a more stable tendency towards the last months of this measurement. In November 2013, the sinkhole received between 3000 and 12000 request per hour.

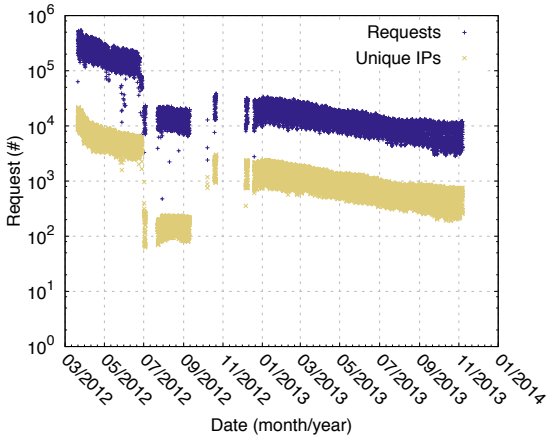


Fig. 5: Total number of requests and IP addresses per hour

Similarly, Figure 6 shows the amount of new IPs per hour over the dataset. If we exclude the initial phase, where the faster decrease is due to the beginning of our measurement period, Figure 6 indicates that the number of new IPs contacting the sinkhole decreases almost exponentially. For example, from the first week of January 2013 to the first week of November 2013, the number of new IP addresses per day decreases from an average of 123 to an average of 48. The number of new IP addresses can indicate both the presence of newly infected hosts, as well as old infections that may belong to networks with dynamic IP allocation (DHCP). In Table 4 we show the number of new Autonomous Systems that appear in the dataset, grouped per quarters. Please note that the first quarter of 2012 only covers the days from the 21st of March to the 31st of March, and the last quarter of 2013 covers the first of October to the 7th of November. Table 4 clearly indicates, for the all duration of the considered dataset, IPs belonging to new ASes have progressively contacted the sinkhole, therefore supporting the theory that the Kelihos.B botnet is still evolving.

As shown in Figure 7 three periods can be identified in which the distribution of job requests is around the same percentage. From the 21st of March until the 29th of June 2012 this average percentage is 13.7%, while from the 20th of December 2012 this average percentage is 32.0%. In the period between the 29th

Year	Quarter	Number of new ASes
2012	1*	4628
2012	2	2805
2012	3	5
2012	4	62
2013	1	69
2013	2	33
2013	3	21
2013	4*	6

Table 4: Number of new autonomous systems per quarter of the year

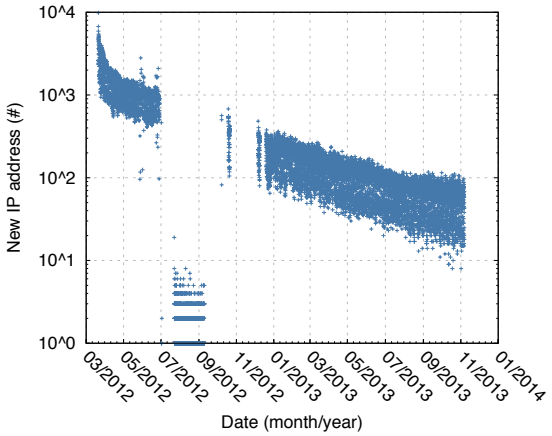


Fig. 6: Number of new IP-addresses per hour over time

of June 2012 and the 20th of December 2012 there are some gaps due to the earlier described corrupted data. Except for those gaps there is a remarkably high average percentage of job requests of 64.1%.

In Figure 8 we have investigated the possible relationship between the requests and the continent where they originate. Figure 8a presents the average number of requests per hour for the continent of origin of a request, normalized by the average number of requests per continents. Similarly, Figure 8b shows the average number of IP-addresses per hour per continent. The normalization allows a direct comparison of the temporal trend of the requests and IP address per hour between different continents. A value of 100% indicates a number of requests or IP addresses equal to the average number. The times in the figures correspond to the times in the dataset as described in Section 3. Although a partition of the dataset in continents is only a simplification of approximating time zones, the results in Figure 8 already indicates that each continent shows a clear day/night pattern. This seems to suggest that the population of infected hosts in the dataset is biased towards laptop and workstations. This observation

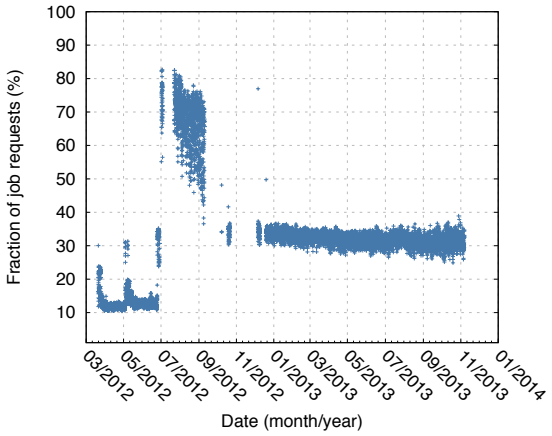


Fig. 7: Percentage of job requests per hour

is also supported by the fact that 87.2% of the infected hosts run Windows XP, as shown in Table 3b.

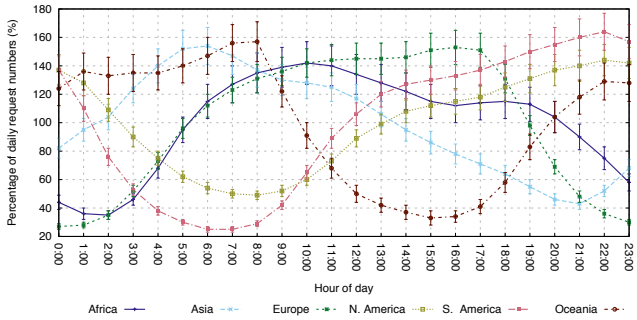
Finally, we have investigated the distribution of the bot version and the operative system of the infected hosts over time. In Figure 9 the development of the version distribution can be seen. All versions follow the exponential decrease of the total number of requests. Interestingly enough, all versions were already present at the time of the sinkholing, and they are persistent over the entire dataset. Also, they percentage with respect to the number of request per hour tends to remain stable. The only notable exceptions are versions 126 and 128, which seem to fade faster just after the sinkholing.

There is no significant change in the distribution of operating systems over time. However, there seems to be a small increase in the share of operating systems running Windows XP. This can be seen in Figure 10.

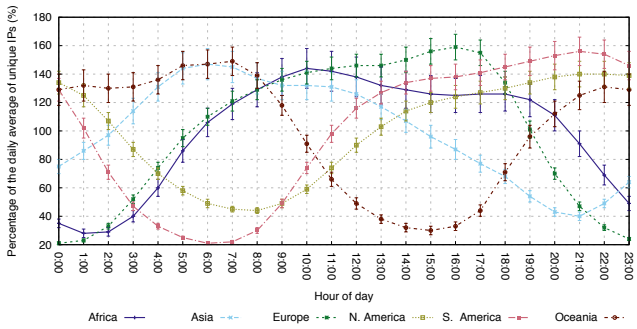
5 Conclusions

In this paper we have investigated the peer-to-peer botnet Kelihos.B. The botnet was sinkholed in early 2012. Using the requests to the sinkholing server we were able to determine what the characteristics of this botnet are. This was accomplished by first analyzing the overall behavior of the botnet in a time-independent fashion and then analyzing the behavior over time to obtain knowledge about how the botnet evolved over time.

In overall, most requests and IP-addresses originate from Poland, although the distribution of IP addresses per country indicate that there is not a one-to-one match between the number of requests and the number of active IP addresses. A telling example is the United States, which is the second most active country per number of requests, but rank only in 10th position per number of



(a) Average number of requests per hour of the day normalised by the average number of requests per continent



(b) Average number of IP-addresses per hour of the day per continent normalised by the average number op IP-addresses per continent

Fig. 8: Geographical request and host distribution per hour of day

active hosts. Overall, 90% of the other IP-addresses have sent less than 135 requests, but the distribution of the number of requests per IP address has a long tail, and a single IP address, located in Montenegro, was able to generate more than 2M requests. Furthermore the AS5617 TPNET Telekomunikacja Polska S.A., which is located in Poland, is the autonomous system from which most requests originate. Finally most bots run version 126 of the botnet software and most bots run on Windows XP.

With respect to the temporal botnet behavior, we can see that the number of requests shows a decreasing trend, although it recently tends to stabilize. On the other hand, the number of new IP addresses contacting the sinkhole seems to constantly decrease over time. The analysis of the requests per continent shows that hosts in different continent follows a day-night pattern. Combined with the observations relative to the operative system of the bots, we can confirm that Kelihos.B primarily targeted personal computer. When considering the distribution of the bot versions and operating systems, our analysis shows that the distri-

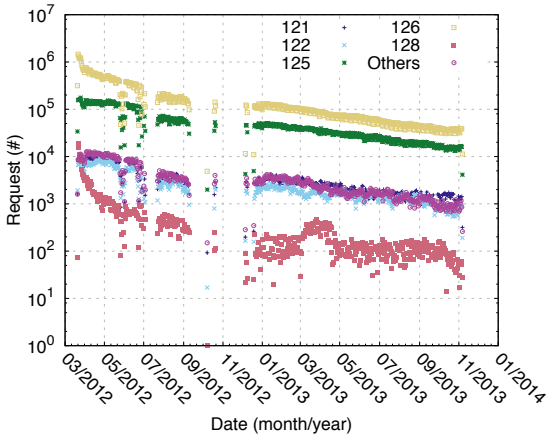


Fig. 9: Number of requests from version number of bots per day

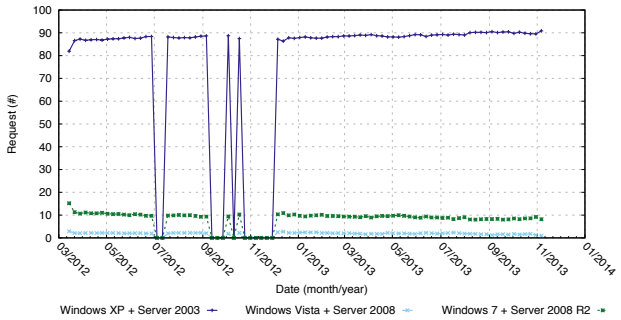


Fig. 10: Percentage of used operating systems for requests per week

butions over time remain the same for the duration of the monitoring period. Also, our results indicate that the bots are still fairly active and the combined analysis of newly appearing IP addresses from different ASes indicates that the infection vector is sill active.

Acknowledgments The authors would like to thank Rogier Spoor (SURFnet) for providing the data used in this research and information about the sinkholing. This research is funded by FLAMINGO, a Network of Excellence project (318488) supported by the European Commission under its Seventh Framework Programme.

References

1. Elliott, C.: Botnets: To what extent are they a threat to information security? Information Security Technical Report **15**(3) (2010) 79–103

2. Ortloff, S.: FAQ: Disabling the new Hlux/Kelihos Botnet. http://www.securelist.com/en/blog/208193438/FAQ_Disabling_the_new_Hlux_Kelihos_Botnet Accessed April 2014 (2012)
3. Kerkers, M.: Characterisation of the Kelihos.B Botnet. In: 20th Twente Student Conference on IT, University of Twente (2014)
4. Werner, T.: P2P Botnet Kelihos.B with 100.000 Nodes Sinkholed. http://www.crowdstrike.com/blog/p2p-botnet-kelihosb-100000-nodes-sinkholed/index.html Accessed April 2014 (2012)
5. Stefan Ortloff : Sinkholing the Hlux/Kelihos botnet - what happened? https://www.securelist.com/en/blog/208214147/Sinkholing_the_Hlux_Kelihos_botnet_what_happened Accessed April 2014 (August 2013)
6. Binsalleeh, H., Ormerod, T., Boukhtouta, A., Sinha, P., Youssef, A., Debbabi, M., Wang, L.: On the analysis of the Zeus botnet crimeware toolkit. In: 8th Annual International Conference on Privacy Security and Trust (PST), IEEE (2010) 31–38
7. Rossow, C., Andriesse, D., Werner, T., Stone-Gross, B., Plohmann, D., Dietrich, C.J., Bos, H.: SoK: P2PWNED-Modeling and Evaluating the Resilience of Peer-to-Peer Botnets. In: IEEE Symposium on Security and Privacy (SP), IEEE (2013) 97–111
8. Werner, T.: Botnet Shutdown Success Story: How Kaspersky Lab Disabled the Hlux/Kelihos Botnet. http://www.securelist.com/en/blog/208193137/Botnet_Shutdown_Success_Story_How_Kaspersky_Lab_Disabled_the_Hlux_Kelihos_Botnet Accessed April 2014 (2011)
9. Knowles, R., Stevens, A.: How Kaspersky Lab and CrowdStrike Dismantled the Second Hlux/Kelihos Botnet: Success Story. http://www.kaspersky.com/about/news/virus/2012/How_Kaspersky_Lab_and_CrowdStrike_Dismantled_the_Second_Hlux_Kelihos_Botnet_Success_Story Accessed April 2014 (2012)
10. Aviv Raff: Kelihos.B is still live and social. https://www.seculert.com/blog/2012/03/kelihosb-is-still-live-and-social.html Accessed April 2014 (March 2012)
11. Alexander Adamov: A Modification of Kelihos Looks for Passwords Stored in Internet Browsers. http://www.lavasoft.com/mylavasoft/malware-descriptions/blog/a-modification-of-kelihos-looks-for-passwords-stored-in-internet-browsers Accessed April 2014 (March 2013)
12. Alexander Adamov: Update on Kelihos Botnet (August 2013). http://www.lavasoft.com/mylavasoft/malware-descriptions/blog/update-on-kelihos-botnet-august-2013 Accessed April 2014 (August 2013)
13. MaxMind: MaxMind GeoIP Database. http://www.maxmind.com/en/geolocation_landing Accessed April 2014 (2013)
14. Ritz, R.: incf.countryutils. https://pypi.python.org/pypi/incf.countryutils Accessed April 2014 (2009)
15. Asghari, H.: PyASN 1.2. https://code.google.com/p/pyasn/downloads/detail?name=PyASN-1.2.zip Accessed April 2014 (March 2010)