# Group 1

# Mirai Botnet: Security Metrics

Romy Bergman *4369408*, Saskia Kooijman *4473795*
Jochem van de Laarschot *4227530*, Giulio Pellizzari *5143160*

23-09-2019

TU Delft
Delft University of Technology

# 1. Introduction and scope

In August 2016, white-hat security researchers discovered the 'Mirai' malware (Paganini, 2016). This malware infects in particular Internet of Things (IoT) devices that are weakly configured and connected to the internet. Examples of these IoT-devices are IP cameras, routers, and printers (Antonakakis et al., 2017). The malware tries to access these vulnerable IoT-devices through a dictionary attack that exploits a set of the most common default credentials. Once access is gained, malware notifies the discovery to the command and control server and then the Mirai malware is installed on the vulnerable IoT-device. The newly infected device then starts searching for other vulnerable devices and repeats the previously mentioned operations on these devices. This way, the Mirai malware spreads quickly creating a botnet (Kolias, Kambourakis, Stavrou & Voas, 2017). After the discovery of Mirai, the source code was published on 'hackforums.net', resulting in fast propagation of the Mirai malware and the instantiation of look-alikes (Antonakakis et al., 2017).

The huge spread of this malware poses a severe concern in terms of availability of internet services, because these infected IoT devices can be used to bring down some of the most important internet service (e.g. DNS). To reduce the threat that Mirai poses to the availability of internet services, the end-users can be informed to create awareness and influence their behavior. However, van Eeten, Bauer, Asghari, & Rand (2010) found informing end-users is not always sufficient. Their research does show how Internet Service Providers (ISPs) are "critical control points" in botnet mitigation. For this reason, ISPs are the entities in the best position to face the issue of the spreading Mirai malware among IoT devices. The ISPs manage every connection in their infrastructure, and therefore have the knowledge about who is sending specific traffic through the network. This knowledge can help them to prevent the spread of the Mirai malware and protect the value of internet service availability. The aim of this report is to propose various metrics which ISPs can use to identify the creation of botnets and prevent them from growing.

In summary, the concept of security (Baldwin, 1997) regarded in this paper is preventing the spread of Mirai malware through IoT devices (*"from what"*) by ISPs ("*whose"*) to secure availability of major internet services (*"which values"*).

# 2. Ideal metrics

In this chapter, ideal metrics are drafted to address the previously names security concept. The ideal metrics give insight about the spread of Mirai malware through the network to help ISPs in mitigating botnets. The metrics are considered ideal, because the formulated information or skills to collect the proper data for these metrics are currently not available to ISPs.

**Botnet size:** *a real time percentage of infected devices relatively to the total number of devices using the network of an ISP*.
When an ISP knows the percentage of infected devices in its network, it can evaluate if, or which, countermeasures have to be taken. Note that currently, numbers on infected devices

making use of a network are possibly over *or* underestimated. This is due to DHCP churning or NAT (van Eeten et al., 2016): these concepts are further elaborated on in chapter 3. When an accurate metric on botnet size is available, it can be regarded whether the botnet may threaten the availability of major internet services in case of an attack. Because the metric is real time, the effect of measures to stop the spread of the botnet can be evaluated immediately.

**End-users:** *the demographics of the owners of infected or vulnerable devices in real time.*
This metric will show the ISPs an overview about the category of people that possess infected or vulnerable devices (e.g. non-educated people, institutions, private companies, etc.). This metric provides ISPs insights about specific users that have bad security practices. For example, it may show that most vulnerable devices belong to poorly educated farmers. Using this metric, the ISP can target this end-user group specifically with information and awareness campaigns.

**Potential losses:** *a real time prediction of potential losses in case all infected devices are used in an attack* and *a prediction of the monetary losses in case all vulnerable devices are infected and used in an attack as well*.

Van Eeten and Brauer (2008) set out a number of potential costs to be incurred by ISPs in case their network traffic consists of a significant amount of malicious traffic. These include: acting on abuse complaints, customer calls and complaints due to *blacklisting* or non-availability of web services, bandwidth costs and brand damage. If a metric could capture all these costs and translate them to potential monetary losses, these costs could be compared with additional costs from countermeasures, such as equipment costs and brand damage (in case the ISP is accused of being too stringent with its measures).

When potential monetary losses are known, an ISP can estimate whether potential investments in botnet mitigation are proportionate to the estimated losses. Although ISPs countermeasures are only effective if done collectively, this metric does underline the importance of taking action and provides incentives for ISPs to act and to coordinate themselves to do this collectively.

# 3. Metrics in Practice

This chapter introduces metrics used in practice that can help an ISP in estimating and mitigating the growth of a Mirai botnet. These metrics are based on various literature sources and are selected since they're applicable to aid in preventing the spread of Mirai malware through IoT devices.

**ISP effectiveness in mitigating botnets**
Asghari et al. (2015) computed botnet infection rates to determine whether countermeasures of ISPs are effective. The metric was developed by mapping IP addresses of infected devices to ISPs and normalizing the amount of infected devices per ISP with regard to the amount of subscribers of each ISP.

**Amount of infected devices per region**

IP addresses can be used to locate *where* most devices are infected. As indicated by van Eeten et al. (2016) the normalized count of infected devices is a known reputation metric per country, ASN and ISP. However, one must note that a simple count of IP addresses can be biased since dynamic IP addresses can overrepresent infected devices. In some cases, one device can be assigned multiple IP addresses over time. The rate in which this happens, is referred to as *DHCP churning* (Moura, et al., 2015). At the same time, because of Network Address Translation (NAT), one IP address could be assigned to multiple devices. This causes underrepresentation of the amount of infected devices. Van Eeten et al. (2016) suggest to aggregate the amount of unique IP addresses over a long period of time (e.g. months) to compensate for these measurement issues.

**Behavior of botnets**

Cetin et al. (2019) analyze the behavior of botnets using location, type of device (router, printer, etc.), protocol (inferred from port number) and the type of network. This data, gathered in collaboration with an ISP, has been used to identify dynamic behaviour of Mirai malware over time. In their study, the information on the behavior of the botnet shows insights in the effectiveness of different means to notify end-users that they have been infected by Mirai and to help them along the remediation process.

The previously identified ideal and practice metrics are evaluated in table 1. The security metrics can be mapped to four different types of metric. These are controls, vulnerabilities, incidents and (prevented) losses (van Eeten, 2015). Controls metrics check the existing implemented measures against a framework of supposed required measures. Vulnerabilities metrics evaluate how security measures perform. Incidents metrics come into use when an attack happened and security is compromised. Finally, losses quantify the consequences of an incident. The goal is to aid ISPs in mitigating botnets. Since the dataset contains data of Mirai incidents and the detection of vulnerabilities in the network aids ISPs to efficiently implement countermeasures, the developed metrics in chapter 5 mainly focus on the vulnerabilities and incidents classifications.

Table 1: Metrics Classification

| **Metric** | **Ideal or in practice** | **Classification** |
|---|---|---|
| Real time percentage of infected devices relatively to the total number of devices using the network of an ISP. | Ideal | Vulnerability |
| Real time demographics of the owners of infected or vulnerable devices. | Ideal | Control |
| Real time prediction of potential losses due to an attack from infected devices and from both vulnerable and infected devices. | Ideal | Losses |
| ISP effectiveness in mitigating botnets. | Asghari et al.(2015) | Incidents |
| Amount of infected devices per region. | van Eeten et al. (2016) | Incidents |
| Behavior of botnets: location, type of device, protocol and type of network. | Cetin et al. (2019) | Incidents |

# 4. Data Description

The provided dataset shows data stemming from a honeypot. This is a system which can detect cyber attacks as it lures Mirai infected devices to connect to the masqueraded vulnerable devices (Antonakaksis et al., 2017). The dataset consists of three different columns. First, the data and time are shown in column one. This indicates when an infected device tries to connect with the honeypot. Second, the IP-addresses of the infected IoT devices are shown. Third, the ports used for scanning for infections are shown. The dataset contains 5947143 entries in total spanned over a timeframe from the 2nd of August 2016 until the 24th of January 2017.

# 5. Developed metrics

Taking the security issue (chapter 1), the ideal metrics (chapter 2), metrics used in practice (chapter 3) and the data at hand (chapter 4) into account, this chapter discusses which metrics could be used for further analysis.
Not all metrics in table 1 can be used because some aspects of these metrics are not measurable or are not available in the dataset. For example, real time measuring is not possible with this provided dataset. Also, the number of subscribers and the type of network are not available to compute the proposed metrics. Therefore, two new metrics are developed and described below.

**Infection rate:** *Percentage of infected devices relatively to the total number of devices for each country per month.*
This metric provides information about the infection rate in the different countries. This is useful, from the ISPs point of view, to identify in which countries Mirai infection is increasing or decreasing. If an ISP is active in highly infected areas, they know it's necessary to take action in order to slow down and stop the malware propagation. Since measurement issues have to be compensated for, results are aggregated over months.

This metric will be computed using the number of infected IP addresses per country per day. As per van Eeten et al. (2016) this result will be averaged over months to compensate for NAT and DHCP churning biases. This number will be normalized with regard to the total number of IoT devices per country. The total number of IoT devices will be estimated using (1) the number of IoT devices in 2016 and (2) the number of IP addresses assigned to each country. The number of IoT devices is based on an IoT report (Ericsson, 2016). According to this report, 35% of the devices is an IoT device. This number will be used to estimate the number of IoT devices per country: 35% of all assigned IP addresses per country is assumed to equal the number of IoT devices for each country. Starting from these two numbers, the average infected devices per month per country and the total number of IoT devices per country, the percentage of infections relatively to the total number of devices will be computed.

Next, to remove ‘noise’ from small countries (few assigned IP addresses), only countries with at least 200 infected devices will be considered. Lastly, only the top 10 countries with the highest average infection rate will be visualized for easier interpretation. To assess the quality of the developed metric, its S.M.A.R.T. properties are regarded in table 2.

Table 2: SMART characteristics of size of infected devices

| % Infected devices per country over time | |
|---|---|
| S | The goal of measuring the percentage of infected devices over time is to get an overview of where in the world the Mirai malware is (mostly) active. |
| M | This metric is measurable, because the number of infected IP addresses is shown in the data. This can then be used to calculate the percentage of infected devices per country over time. |
| A | The calculation of this metric can be performed with the means available, because the infection of IoT devices has been measured. The result of this is shown in the data set used to compute this metric. |
| R | This metric provides necessary information for the ISP, because it shows where in the world the Mirai malware has been most active during the time of analysis. Based on this, the ISP can take suitable action in allocating its resources in fighting the spread of Mirai malware. |
| T | The percentage of infected devices per country will be aggregated over a timespan of a month. |

Nevertheless, this metric has some limitations. Firstly, the number of IP addresses is averaged per month due to the issues described by van Eeten et al. (2016). Therefore the metric can not provide a timely (e.g. real time) description of the infection rates per country. Secondly, the IP addresses are mapped to countries, instead of ISPs. As ‘metrics in practice’ show, these can be be further refined by mapping them to ASNs and ISPs to better fit the ISP perspective that has been elaborated upon in the previous chapters. Finally, a significant limitation is the rough estimation of the number of IoT devices belonging to each country. This methodology contributes to a less reliable metric, but was necessary due to informational constraints.

**Behavior of Mirai botnets:** *Port distribution over time*

This metric shows which ports are used by the Mirai virus to search for vulnerable devices. Because the information is shown over time, the dynamic behavior of Mirai can be regarded. With insights provided by this metric, ISPs can adapt their countermeasures to the behavior dynamics of Mirai.

The metric will be computed by counting the occurrences of the ports scanned during the period of observation of our dataset. To be consistent with the previous metric, the port count will be grouped by month in order to provide the same scale of comparison. To assess the quality of the developed metric, its S.M.A.R.T. properties are regarded in table 3.

Table 3: SMART characteristics of behavior of Mirai botnets

| Port distribution over time | |
|---|---|
| S | The goal of measuring the port distribution over time is to track what ports are being scanned by the Mirai malware. |
| M | An exact frequency of accessed ports per month is shown. Showing the dynamic behavior over time makes this metric measurable. |
| A | Data about scanned ports has been measured using the honeypot. The port distribution can be easily computed by means of statistical analysis. Therefore, this metric is not paired with high costs, which makes it actionable. |
| R | The port distribution is highly relevant as it can give the ISP directions on the exact destination of Mirai traffic. This information gives ISPs the means to know what ports to possible block in order to mitigate Mirai botnets. |
| T | The data is aggregated over months. Hence, each month the result can be regarded as a milestone. If the data indicates specific ports being scanned by Mirai, the ISPs know what countermeasures to take for their infrastructure. |

However, the metric has also some limitations. It does not show the development of the virus per country. Nevertheless, ISPs can still regard the dynamic behavior of Mirai and adapt their mitigation strategies to this behavior. The bias issues are also still present in this metric. Due to time constraints during the data analysis the data will be grouped by month, instead of taking the average over months of the count per day.

# 6. Data Analysis

Using the metrics developed in chapter 5, the dataset originating from the IoT honeypot is analyzed.

**Percentage of infected devices per month per country**

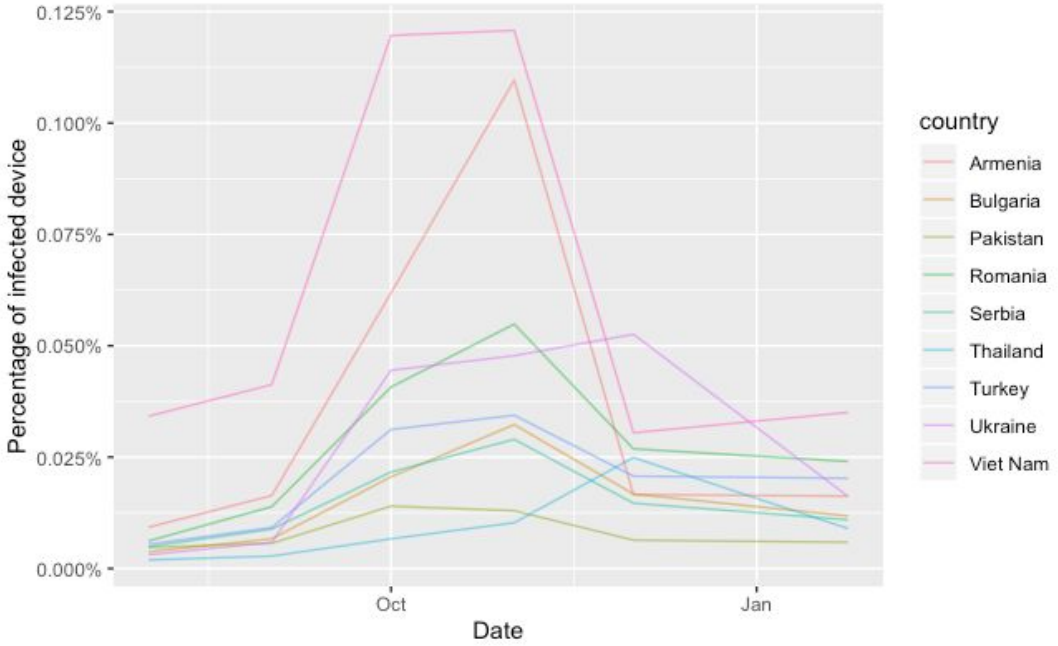


Figure 1: line graph size of infected devices per country

Figure 1 shows the percentage of infected devices per country. Notable is that all countries in this dataset are having an overall increase in infection percentages around the beginning of October 2016. This can be related to the fact that on September 30th, the Mirai malware code was published online (Antonakakis et al., 2017). In the graph a large growth of relative botnet size in Vietnam is evident and confirms what is underlined by Antonakakis et al. (2017). These may be signs that the mitigation strategies of ISPs in countries such as Vietnam and Armenia are worse than mitigation strategies of other ISPs or evidence that their investments in mitigating this threat are still not enough.

**Port distribution over time**

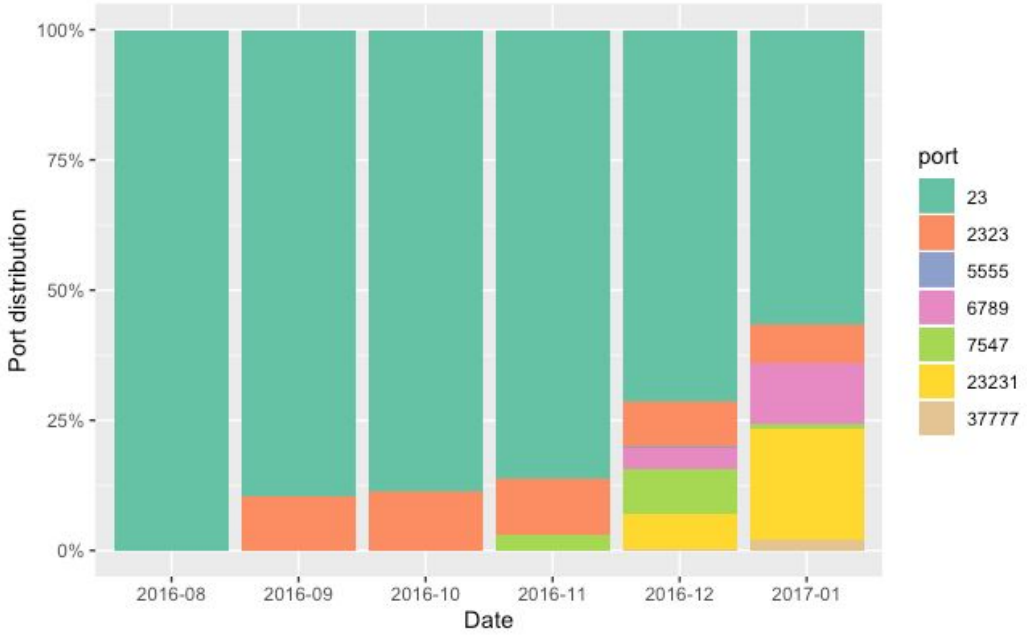


Figure 2: stacked bar chart port distribution over time

Figure 2 shows the stacked bar chart of the cumulative port distribution over time. As of January 2017, over 50% of the scanned ports were port 23. However, as time evolved additional ports were being scanned. This is due to the publishment of the code that causes the proliferation of new Mirai versions applying the same attack pattern but on different ports (Antonakakis et al., 2017). Therefore, monitoring the scanning behavior of the ports aids ISPs in implementing their countermeasure dynamically. More Mirai versions can emerge and specific Mirai botnets may grow faster than others, causing the need for ISPs to act on multiple ports and change their strategy over time.

# 7. Conclusion

This paper presents two metrics to inform ISPs on the spread of Mirai malware through IoT devices. By being informed on this spread, ISPs can effectively design countermeasures and evaluate their effectiveness. When ISPs design effective countermeasures, the availability of major internet services can be safeguarded. This paper shows that measurement issues caused by NAT and DHCP churning heavily influences the metrics used in practice and developed in our research. The developed metrics are designed using insights from an ideal

world, where all information is instantly available, metrics used in academic literature and the dataset at hand.

As a result, one metric describes the infection rate by showing a percentage of the infected devices relatively to the total number of devices for each country. To compensate for bias issues, results are averaged over a month. This reduces timeliness of the metric, but increases accuracy. By analyzing the data using this metric, it is found that ISPs in the countries Armenia and Vietnam should increase their efforts in slowing down the spread of the Mirai botnet. It must be noted that the methodology of estimating the number of IoT devices per country is rather rough. This might influence the results.

The other developed metric describes the behavior of Mirai bots by showing the port distribution over time. The metric shows the dynamic behavior of the botnet. With this insight, ISPs can adapt their countermeasures to this behavior. The analysis shows that the virus evolves over time. This is probably due to the release of the code, which resulted in adaptations of the original virus.

The developed metrics can be a step to motivate ISPs to be an active player in mitigating botnets. Both the percentage of infected devices and the port distribution can be used as insights to initiate appropriate countermeasures. By using incident data, network vulnerabilities are detected and Mirai botnets can be mitigated.

# References


Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., ... & Kumar, D. (2017). Understanding the mirai botnet. In 26th {USENIX} Security Symposium (pp. 1093-1110).

Asghari, A., van Eeten  M. J. G. & Bauer J.M., "Economics of Fighting Botnets: Lessons from a Decade of Mitigation," in IEEE Security & Privacy, vol. 13, no. 5, pp. 16-23, Sept.-Oct. 2015.

Baldwin, D. A. (1997). The concept of security. Review of international studies, 23(1), 5-26.

Cetin, O., Hernandez Ganan, C., Altena, L., Kasama, T., Inoue, D., Tamiya, K., ... van Eeten, M. (2019). Cleaning Up the Internet of Evil Things: Real-World Evidence on ISP and Consumer Efforts to Remove Mirai. In Network and Distributed System Security Symposium (NDSS) 2019

Van Eeten, M. J. G. & Bauer J.M., (2008). Economics of Malware: Security Decisions, Incentives, and Externalities.

Van Eeten, M. J. G., Bauer, J. M., Asghari, H., Tabatabaie, S., & Rand, D. (2010). The role of internet service providers in botnet mitigation an empirical analysis based on spam data. TPRC.

Van Eeten, M. J. G., Lone, Q., Moura, G., Asghari, H., & Korczyński, M. (2016). Evaluating the impact of AbuseHUB on Botnet mitigation.

Van Eeten, M. J. G.,. (2015, Januari 25). 2b Measuring Security Levels [YouTube]. Retrieved from: https://www.youtube.com/watch?v=jwVNuWi4EwE

Ericsson (2016). IoT Forecast. Retrieved from https://www.ericsson.com/en/mobility-report/internet-of-things-forecast

Ganan, C. (2015, Januari 25). 2a Measuring Security Levels [YouTube]. Retrieved from: https://www.youtube.com/watch?time_continue=499&v=fvUkyVUBQMY

Kolias, C., Kambourakis, G., Stavrou, A., & Voas, J. (2017). DDoS in the IoT: Mirai and other botnets. Computer, 50(7), 80-84.

Margolis, J., Oh, T. T., Jadhav, S., Kim, Y. H., & Kim, J. N. (2017, July). An In-Depth Analysis of the Mirai Botnet. In 2017 International Conference on Software Security and Assurance (ICSSA) (pp. 6-12). IEEE.

Moura, G. C., Ganán, C., Lone, Q., Poursaied, P., Asghari, H., & van Eeten, M. (2015). How dynamic is the isps address space? towards internet-wide dhcp churn estimation. In 2015 IFIP Networking Conference (IFIP Networking) (pp. 1-9). IEEE

.

Paganini, P (2016). Linux/Mirai ELF, when malware is recycled could be still dangerous. Retrieved from https://securityaffairs.co/wordpress/50929/malware/linux-mirai-elf.html