WM0824TU Group 9

Chris Berg, 4216776
Martin Koster, 4371011
Radinka Yorgova, 4952545
Wilko Meijer, 4224426

September 24, 2018

Code and data:
https://github.com/wkmeijer/WM0824TU_Group9

1 Security Issue

The data from badpackets.net (1) speaks to security issues which are caused by Mirai. Mirai is malware targeting IoT devices with the goal of creating a botnet of IoT devices. IoT devices are infected through other infected IoT devices which are scanning the Internet for vulnerable IP addresses. When an infected IoT device found a vulnerable IP address it uses a table with default factory login passwords to get the Mirai malware on the target device.

There are a few actors relating to this threat. First of all, there is the owner of the IoT device. In general their security issue is only minor, since the malware is only using the processing power of their device, so their loss is mainly slightly slower devices. For large organization with lots of infections the issues are larger, since the attack may be attributed to the organization. This could become a large security issue as victims of attacks by the botnet might retaliate against the organization from which the attack originates. Besides this there is also the potential reputation loss resulting from the attribution. So for them there is incentive to mitigate infections.

The second set of actors are the victims of the botnet attacks performed by the mirai bots. The security issue for these victims consists mainly of disruption of service through DDoS attacks. However, since our dataset has no information about the targets of the botnet, these actors are mostly out of scope.

Lastly, there are the actors which control the networks over which the attacks are performed, the ISPs and other owners of Autonomous Systems. The security issue for these actors is flooding of their network causing a disruption in service, however most ISPs are well equipped to handle peaks in network activity. When an ISP is offering other services, the botnet might pose a security issue in other ways as well. For example spam e-mails sent by the botnet. However, even though their security issue might be minor, they are very well equipped to mitigate the security issues of the other actors. ISPs have the power to control the network flows and can block the spreading of the malware (2) and block infected hosts. Therefore we will also focus on these actors in this report.

The original Mirai code is available to the public (3). So now, after the creators stopped, there are still attackers who use this code or even evolve this code. The security issue is therefore still relevant.

2 Methodology

To understand the topics of the assignment everybody watched the video lectures and read the papers. In this way it was a lot clearer for us to know what we can expect and use from the metrics. After that we wanted to first completely understand what the problem is with the subject of our group. With this information we could answer also the first question. For the other questions we could use information from the lectures. Everyone worked separately on the following tasks:

- Extracting data
- Ideal metric
- Existing metrics
- Metrics for dataset

After all this information is collected we can actually apply the metrics on our dataset and evaluate the results.

3 Metrics

Security metrics are critical tools designed to facilitate decision-making and improve performance and reliability through collection, analysis, and reporting of relevant performance-related data. The right metrics should help the decision maker to allocate security spending and justify the investments, to get actual security and gain certain benefits.

For a decision maker to justify any security investment there are two conditions: to quantify the costs of security and to quantify the benefit. The connection between cost and benefit is described by the security production function (4).

This function goes through a middle step security level. First it maps the cost of security into the security level and then in the second step, the security level into the benefit.

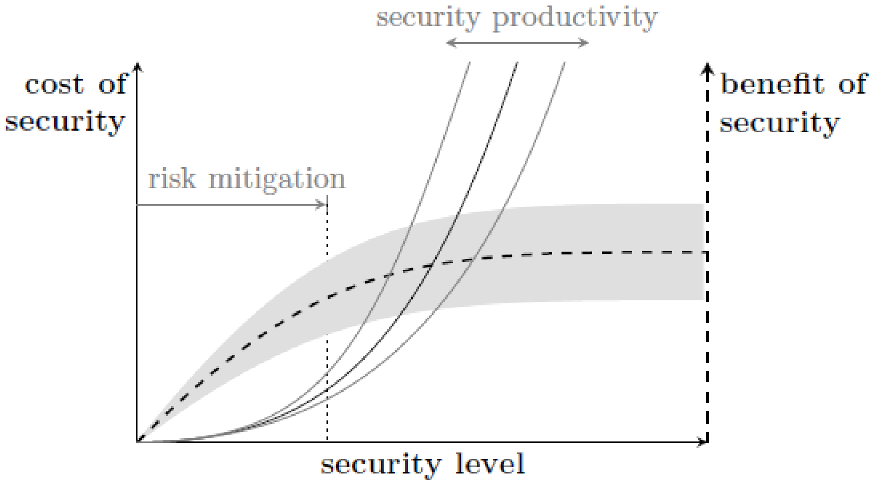

Figure 1: Security production function.

The graphic in Figure 1 (4) shows that from some point spending more and more will get you smaller and smaller improvement in security level and also, after a certain point, further increasing the security level leads only to marginal benefits.

But all that is in the perfect theoretical model.

3.1 Ideal metrics for security decision makers

Before we answer to the question what is the ideal security metric in our study case we provide a short overview of the security metric classification.

There are described four types of metrics for measuring the security levels. Or more precisely said four types of metrics measuring indicators that reflect different aspects of the security level.

The four categories metrics are based on controls, on vulnerabilities, on incidents and on prevented losses.

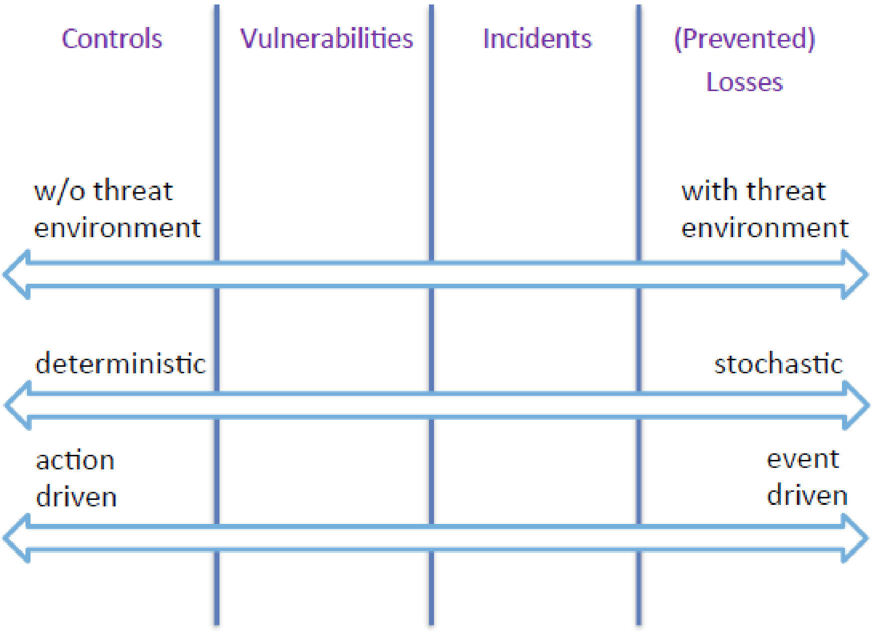

Figure 2: Types of metrics

The first category are metrics based on controls. Controls are the measures implemented to mitigate risk. They can be physical (door locks, building specifications) organizational (incident response team/employee), procedural (credential management) or technical (like firewalls). Controls exclude the threat environment. Vulnerabilities are weaknesses in the controls that can fail by a certain type of attack.

The next type metrics is based on vulnerabilities. The vulnerabilities can be known and unknown. For the known there are different approaches to check whether known vulnerabilities are present in the systems. Against the unknown vulnerabilities the typical techniques are penetration testing or red teaming. It is important to stress that the vulnerability metrics are driven by attack scenarios and they are static.

The last two metrics are driven by the incident itself. The very last metric, prevented losses, try to map incidents on losses, i.e. it is driven also by the economic impact of the incident. They are the most advanced metrics.

Incidents and prevented losses are stochastic. They are driven by unknown attacker behavior and the actions of the defender. Moreover they are driven by events (for example a data breach) wheres controls are mainly driven by actions (installing a new firewall).

Back to our Mirai-like Botnet study case we focus on improving the Security Level of the Internet Service Providers. As discussed already, there are few security issues for the ISPs caused by botnets in case of attack:

- malicious traffic that is generated by botnets and carried through their networks
- if the ISPs offer email services - the amount of spam produced by botnets
- infected devices in their network, this can force ISPs to reestablishing the integrity of their networks. An example for this is from November 2016 with Deutsche Telekom patching routers soon after the attack (6).

For our ideal security metric we propose the following list of activities which will lead to increasing the security level of the ISPs. Each of them cost a security investment which has to be approved by the shareholders.

- Prevention: provide proactively end-point security solutions which prevents customer systems to get infected with malware and become part of a botnet. For example provide anti-virus software or secure routers to the customers or inform/educate the customers for the threats imposed by botnets
- Detection: the goal is an infected customer systems to be detected and subsequently isolated as soon as possible, i.e. continuous monitoring of the data traffic and immediate actions.
- Notification: this should be in two directions: inform the infected customer and share the information about the infected system with stakeholders such as peer ISPs. This will improve the prediction of next attempt from a bot.
- Remediation and recovery : provide a removal malicious software to the infected customer; provide information to the customer about possible effects of recovery on personal data and accounts.

- Analysis 1: analyze the shared information for the infected systems and improve the data monitoring with isolating/restricting communications with systems from regions with dense botnet.
- Analysis 2: make a characteristic of the customers to estimate the risk of botnes activity- higher usage rates of pirated software are associated with higher botnet activity, while higher education levels, as an indication of technical competence, are associated with lower levels. This study should not violate the privacy law.
- Analysis 3: analyze possible infection with ransomware. Ransomware can be distributed by botnets, and therefore investing in fighting botnets can prevent of eventual losses from ransomware. (Ransomware, is a type of malware that prevents users from accessing their system or personal files and demands ransom payment in order to regain access).

The proposed activities defines our ideal security metric as a combination from metrics based on controls, vulnerability, incidents and prevented loses. The ideal metric should also have reasonable proportion between $cost\ of\ security$- *security level*- $benefit\ of\ security$ given in Figure 1.

Some of the given steps are addresses as recommendations in (7). It is clear that this data base does not equip us to provide full description of the proposed metric. Never the less the data are suitable for an extensive analysis of step 4 - Analysis 1.

3.2 Metrics that exist in practice

As specified in Chapter 1, the victims from a botnet attack we are focussing on are ISPs. In (5) they discuss metrics to detect the bots of a botnet and (hopefully) before a botnet attack. First of all they take the relationship into account. By a relationship is meant: the communication over one protocol. There is assumed that they bots are dense in their topology and the master bot is in the center of that.
Besides relationships, also the response is taken into account. When the master bot sends a task to a bot, the bot will execute a pre-programmed script. The response time is always the same and the action is accurate. The human (legitimate) behaviour in contradiction to this bot behaviour is a lot more variable in response time. At last, the synchronization is an import aspect of botnets. They have assumed that all the bots are synchronized. So because of that all bots will act at the same time when the master sends a task.
With all these three metrics, bots can be traced and hopefully track the master bot to prevent any botnet attacks.

Of course, as ISP company, you also want to be prepared for such an attack. The only thing they can do is to keep their environment secure. This have to be checked with the help of metrics. Earlier we discussed the concept of the security production function (Figure 1). There are existing metrics which measure those three units (cost of security, security level and benefit of security). Those metrics are shown in Figure 3. The metrics for the costs and benefits of security are shown order from concrete to abstract. The metrics more to abstract are harder to measure. The security level is in the range from deterministic to probabilistic. Again if the metrics go to the probabilistic side they are becoming harder to measure.

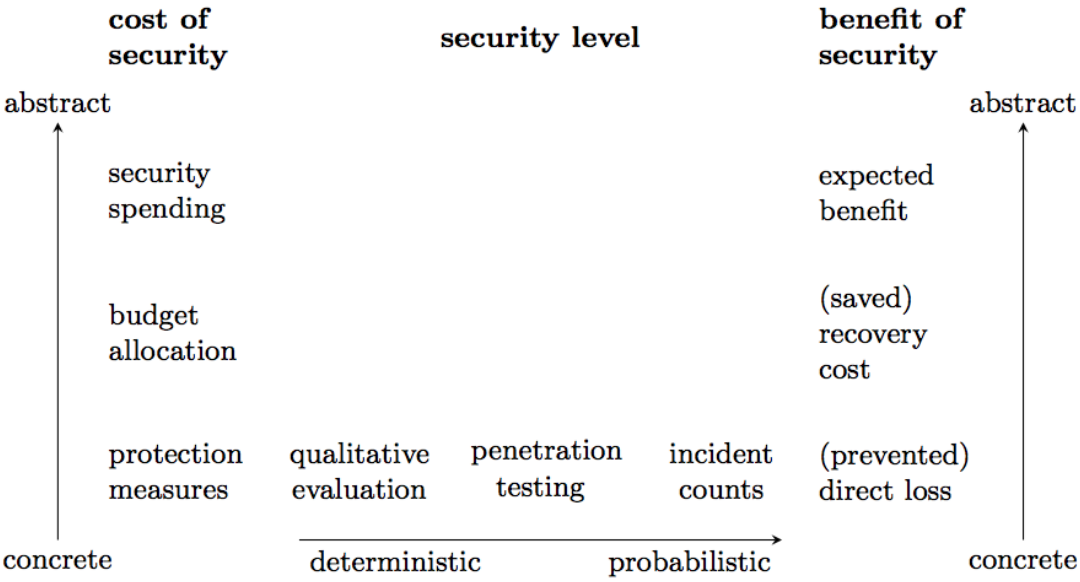

Figure 3: Metrics in the different categories

3.3 Definition of metrics from dataset

The dataset consists of 233 629 entries depicting IP addresses which have shown behavior related to an infection by the Mirai malware. An IP address is added to the dataset the first time an infection is detected, so it doesn't say anything about whether or not the machines behind those IP addresses are currently infected. The metrics can therefore only focus on new infections and thus nothing conclusive can be said about the size of the botnet. The data also contains no information about who is the target of the infected machines. Therefore the metrics will mostly focus on giving insight on how the malware spreads and thus help the security decision makers on the side of infected machines and not the victims of attacks performed by the botnet. The metrics can also show whether mitigations against the Mirai malware actually have an effect on how it spreads.

The dataset contains the following information for each entry:

- IP address - IPv4 address which shows behavior of infection
- Autonomous System (AS) - The name of the subnet to which the IP address, usually the name of the organization that owns the AS
- Country - The country where the AS and IP address is registered
- ASN - Autonomous System Number
- Date First Seen - The time the IP was first seen to have been infected

3.3.1 New infections observed over time

This metric can show at what rate the mirai botnet is spreading. When all entries are binned within a certain time frame, say 2 weeks, it becomes clear how much the amount of new infections changes of time. Note that from this metric no conclusions can be drawn on the size of the botnet, since the data only says something about new infections, not current total infections.

3.3.2 Infections per country normalized on country size

All the entries are grouped per country and normalized on country size. This metric will show whether or not there are large differences between new infections per country. This is useful information for security decision makers in determining the chance of infection. Since countries with a large amount of internet connections will almost inevitably also have more new infections, the data is normalized on the amount of internet connections per country. A better metric would be to also visualize this over time, so the development of infections are clear within a country, however this is very hard to represent. So this could be done for specific countries of interest.

3.3.3 Infections per autonomous system normalized on size over time

All the entries are grouped per AS and plotted over time. This metric is very similar to the one discussed in subsubsection 3.3.2, only now the development of infections per AS is measured. This can give insight in the spread of infections within autonomous systems. Which is again useful information for determining the chance of infection. This metric can also show whether certain ISPs (owners of most of the autonomous systems) have more infections in their network. When the infections are skewed towards certain ISPs this could have consequences for their reputation. The ISPs might be blamed for negligence in fighting back against the botnet, since apparently they are doing something different to non-infected ISPs. Since autonomous systems, like countries, can be very different in size the data is normalized on the size of the AS.

3.3.4 Amount of unique countries getting new infections over time

The data is sliced into pieces of equal length in time, and for each of the slices the amount of unique countries is plotted. This metric measures how many unique countries are infected within a certain time period, say a week. It shows whether the malware is focused only on a few countries, or that the spread is fairly random and is correlated with the total number of infections as shown with subsubsection 3.3.1.

3.3.5 Amount of unique autonomous systems getting new infections over time

Metric with the same goal as subsubsection 3.3.4, only now for autonomous systems.

4 Evaluation of metrics from dataset

It is hard to say anything meaningful about the quality of the data, because not much is known about how the data is collected. Network traffic is sniffed for a fingerprint left by the Mirai malware, however for all the practical reasons the dataset can never be a complete list of all infected devices. A major problem is therefore determining what the bias is of the data. We will work under the assumption that there is minimal bias.

4.1 New infections observed over time

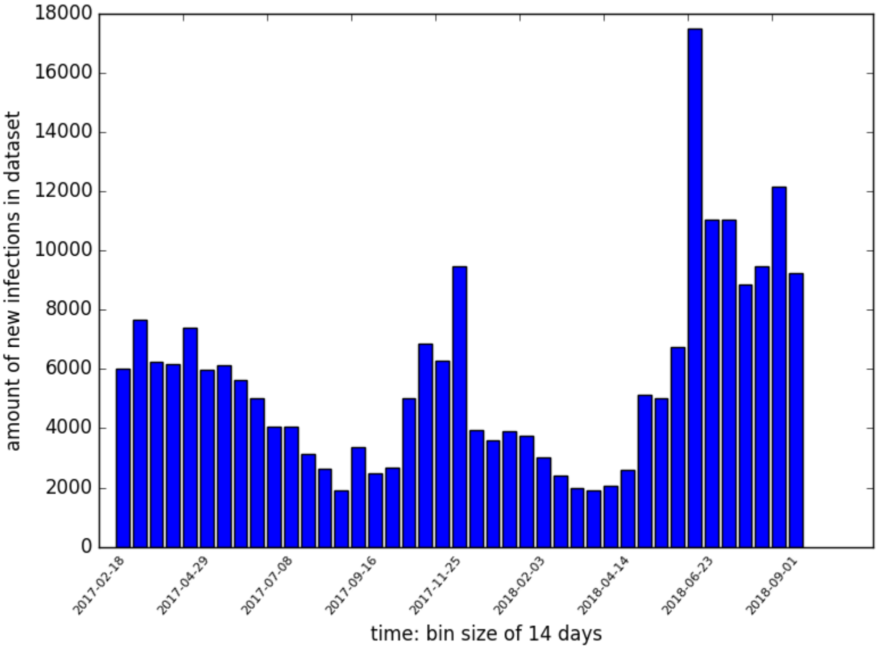

Figure 4: Amount of new infections over time

When observing the rate at which the Mirai botnet is spreading (Figure 4), a waveform seems to emerge. The waveform might be periodically, but because of the short timeframe (2.5 periods) there is absolutely no certainty about that. The increases in spread rate might cohere with the discovery and exploitation of new weaknesses in autonomous systems.

4.2 Infections per country normalized on country size

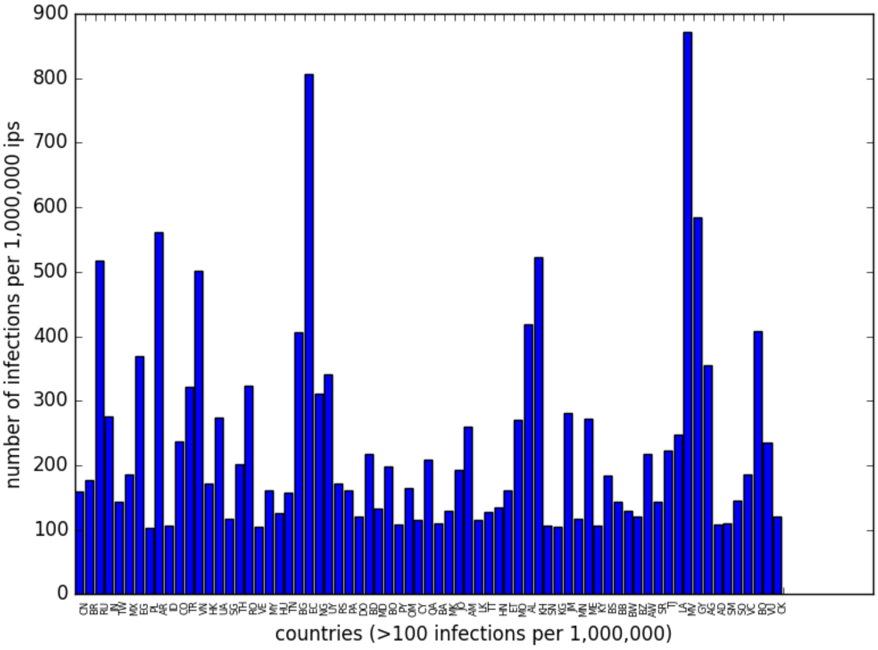

Figure 5: Amount of infections per country

When we plot the amount of infection in the dataset per country, clear differences start to appear. In Figure 5 only the countries with more than 100 infections per 1 000 000 IPs are shown for better visualization. It becomes very clear that certain countries have a much bigger problem with infections than others. Especially Russia, Argentina, Venezuela, Ecuador, Cambodia, Maladives, and Guyana seem to suffer from an abnormal amount of infections. For some countries this can be explained by their low number of internet connections in general, so with very few connections they have a high amount of relative infections. However, for Russia and some of the large South-American countries, with a decent amount of internet connections, their infection numbers are really high. This means that ISPs operating in those countries have a very high chance of having a lot of infected devices in their network. Thus those ISPs should conclude from this graph that they should try to mitigate this, as they stand out against the rest, making them more prone to reputation damage. For organizations residing in any of the countries that have an abnormal amount of infections, this means that security audits are more likely to find infections. Thus they might conclude from this data that it is worth the investment, since the chance of infection pretty high, e.g. 1 in 1000 for Ecuador.

4.3 Infections per autonomous system normalized on size over time

Unfortunately visualizing this metric proved to be too much of a challenge, since there are many unique autonomous systems present in the dataset.

4.4 Amount of unique countries getting new infections over time

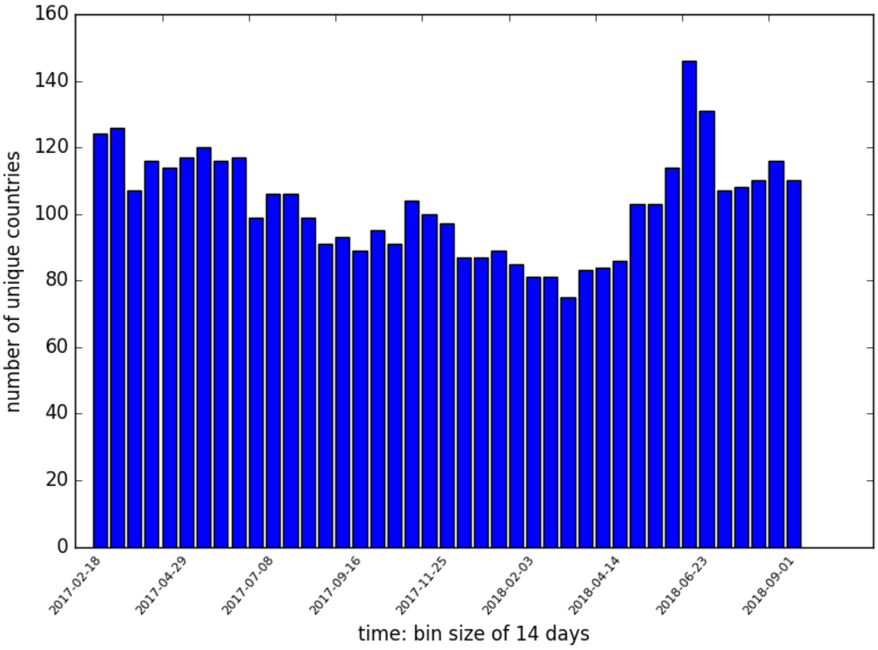

Figure 6: Amount of unique countries getting new infections over time

When observing the amount of unique countries getting new infections (Figure 6), a downwards trend until summer 2018, and an upwards trend from there on becomes visible. When compared to the spread of the Mirai botnet (Figure 4) and the amount of unique autonomous systems getting new infections (Figure 7) this trend becomes, contrary to the expectations, visible only in the amount of unique autonomous systems getting infections over time, and not in the amount of new infections over time.

4.5 Amount of unique autonomous systems getting new infections over time

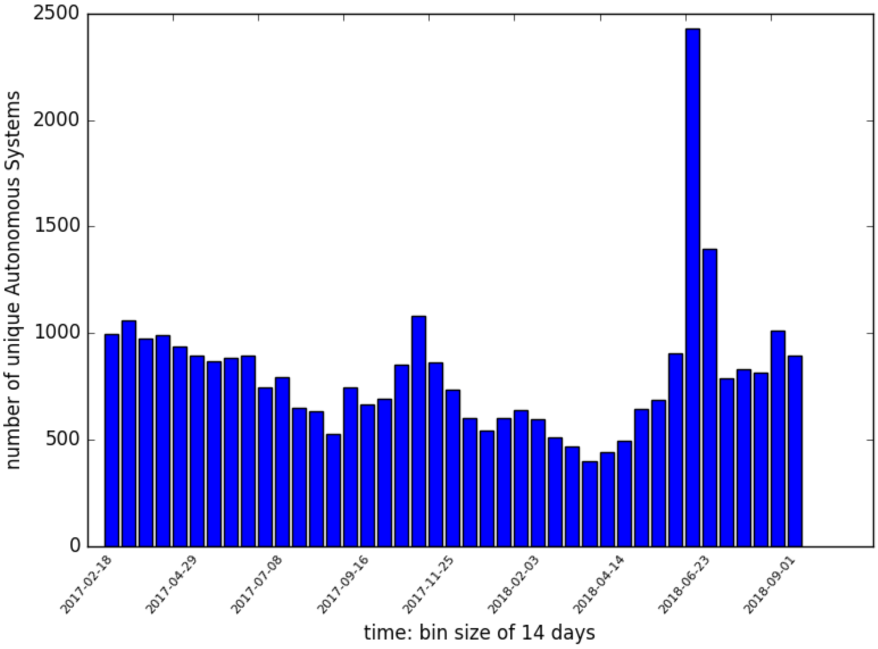

Figure 7: Amount of unique autonomous systems getting new infections over time

When the amount of unique autonomous systems getting new infections (Figure 7) is compared to the spread rate of the Mirai botnet, it becomes clear that the two correlate. The peaks and valleys are the same, from which the hypothesis can be derived that the infection of new autonomous systems results in an increase of the spread of the Mirai botnet.

4.6 Observations from the evaluation of metrics

The metrics give some insight in the chances of an organization, be it an ISP or some other organization has infected devices in the network. Especially on the country level, very clear differences start to emerge. Furthermore the graphs over time give some indication of how the amount of infections are still quite high. This information is useful for security decision makers, because from this they can conclude that the security issues are still relevant, as new infections keep appearing. For a more detailed risk assessment the normalized infection rates for AS's should also be present, so security decision makers can specify the risk for their network specifically.

References

[1] Mirai-like Botnet: Bad Packets Report, https://mirai.badpackets.net/, accessed 13 Sep 2018

[2] Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., Durumeric, Z., Halderman, J.A., Invernizzi, L., Kallitsis, M. and Kumar, D., 2017, August. Understanding the mirai botnet. In USENIX Security Symposium (pp. 1092-1110).

[3] Leaked Mirai Source Code for Research/IoC Development Purposes, https://github.com/jgamblin/Mirai-Source-Code, accessed 18 Sep 2018

[4] R. Böhme,Security Metrics and Security Investment Models, Advances in Information and computer Security, pp 10-24, 2010.

[5] Akiyama, M., Kawamoto, T., Shimamura, M., Yokoyama, T., Kadobayashi, Y., Yamaguchi, S. (2007, January). A proposal of metrics for botnet detection based on its cooperative behavior. In Applications and the Internet Workshops, 2007. SAINT Workshops 2007. International Symposium on (pp. 82-82). IEEE.

[6] Deutsche Telekom. Telekom-hilt. https://www.facebook.com/telekomhilft/photos/a.143615195685585.27512.122768271103611/1199966633383764/?type=&theater.

[7] Pijpker, Jeroen, and Harald Vranken (2016) ”The role of Internet Service Providers in botnet mitigation.” Intelligence and Security Informatics Conference (EISIC), 2016 European. IEEE