Assignment block 2 - Group 9

What security issue does the data speak to?

We identify a security issue caused by a botnet infection, discuss the **actor** from whose point of view we analyse the issue, and finally discuss the three components that make it a security issue (whose security, what to secure, and security from who).

The dataset collects packets originating from the Mirai botnet. A botnet such as Mirai consists of infected computers that are used to infect new computers and perform arbitrary attacks as ordered by the controller of the botnet. Infected machines pose a risk to the security of the network, since they can be used to perform (DDoS) attacks, act as ransomware, steal data, and more.

We will discuss the security issue of infected machines from the perspective of an IT admin who is responsible for the security (confidentiality, integrity, availability) of the network. We will discuss the metrics that are relevant for them when weighing controls against risks.

This is a security issue because:

- *Security of what?* There is a danger to each of the different elements of CIA-security: Confidentiality is threatened because data can be stolen. Integrity is threatened because data on infected machines can be altered. Availability is threatened when a ransomware attack is performed.
- *Security of who?* Upholding the security of the network is one of the main tasks of the IT operator, thus their security is threatened.
- *Security from?* The threat is instigated by the botnet operator, thus there is a specific adversarial actor who causes the issue.

Ideal metrics for security decision makers

The role of a security decision maker is to weigh the cost of a security measure against the increase in safety, and thus the decrease in incidents. The ideal metrics for this role would make it easy to perform such a cost/benefit analysis. We therefore chose a set of metrics that describe the risk and cost of incidents occurring. Because the security issue we discuss is about the Mirai botnet, we specifically chose metrics related to botnets. With an *attack*, we mean an attempt at infecting a machine.

1. Number of attacks against (similar) networks
2. Probability of an attack succeeding
3. Growth of botnet over time
4. Potential damage/cost per machine infection
5. Cost of infection prevention
6. Cost of infection reparation

To get an idea of the risk to security issues, the decision maker wants to know the likelihood of the security issue occurring and the impact of this issue. With these metrics the security decision maker can make an appropriate analysis of the current level of security, assess how new security measures with their individual costs would impact the security level and therefore determine the potential security benefits of any new measures. In the following table we summarize expand upon the metrics above and place them in a context that enables risk analysis for the security issue:

	Likelihood	Impact
Botnet infection Attacks	(1) How often are similar company networks attacked? (2) What is the likelihood an attack is successful given the level of security of a network? (3) How is the botnet changing? Is it becoming more or less capable of damage?	(4) What are the damages associated with a machine infection? (5) What are the costs of preventing a machine from being infected? (6) What are the costs of restoring/repairing a machine?

Knowing the likelihood and impact, and thus the risk, of the security issue, the IT admin can now properly assess possible controls. First they must calculate the *security costs* of the controls, during which metrics 5 and 6 should be applied. Then they must expand on this to find the *level of security* offered by the controls. This should be done in the context of the security issue, thus metrics 1, 2, and 3 should be considered. Finally, they can weigh the level of security against the security costs to find *security benefits*. To do this, the cost of a machine infection must be known, which is explained by metric 4.

Metrics used in practice

In this section we discuss some literature on detecting botnet infections, botnet growth, the behaviour of botnets, and botnet analysis.

Akiyama et al.[2] discuss metrics for detecting that a botnet infection is present in your own network. They focus on three metrics: relationship, response and synchronization, which we will briefly detail. The relationship metric uses the fact that the bots have a bot-master somewhere, which each bot communicates with. By checking network traffic, we can attempt to find a single remote server that many of our devices are communicating with; this could be a botnet master server. The response metric relies on identifying a consistent response time between a bot receiving a signal from its master and its response to that signal. This allows us to distinguish between human-caused traffic and automated traffic. The synchronization metric uses the simultaneous nature of the entire botnet. All (or at least a large portion of) the bots will take the same actions, such as a DDoS attack or reporting their status to the master, at almost the same time. The detection of such activity spikes

could indicate a botnet infection is present. These three metrics can be used to identify compromised systems in a network traffic analysis.

Antonakakis et al.[3] use a combination of techniques for measuring botnet size and tracking the evolution of the botnet's capabilities. The botnet size was measured by making use of a network telescope. Using this network telescope, infected hosts scanning for new victims could be found. The metric used to estimate the size of the botnet is the number of hosts which actively scan for potential infectants at the start of every hour. In order to track the capabilities of the botnet a telnet honeypot was used to obtain infected binaries in combination with a set of binaries from VirusTotal. The total number of unique binaries gives an estimation on the size of the set of capabilities of the mirai like botnet. These could be analyzed further to determine architecture and vendor targets.

Imperva Security[4], an IT security provider, has broken down the behaviour of the Mirai botnet. The results show various elements of the botnet behaviour, some of which can be measured as metrics. First, bots will perform scans on the entire IP-address range, except for a limited set of IPs such as the network of the Department of Defense. The metric can be measured by recording the destinations of outgoing packets. Second, the method used by Mirai bots to infect new hosts is by attempting a known list of username/password combinations. By recording attempted logins, an attack by a Mirai bot can be detected, and the risk of infection increases. The metric would describe the number of attempted logins using credentials from this list.

Jenkins [5] discuss the use of network telescopes such as the UCSD Network Telescope [6] to detect and visualize the growth of botnets such as Mirai. The network telescope measures random scan efforts by botnets to find new infectable hosts. As long as the botnet keeps searching for new hosts to infect, the rate of scan packets over time will therefore increase as the botnet size increases. The network telescope did indeed show a big increase in telnet activity starting in June 2016, matching Mirai's growth.

In order to detect IRC based botnets Karasaridis et al. [7] have created an anomaly-based passive analysis algorithm that detects IRC botnet controllers with a 2% false positive rate, without relying on known botnet signatures or binaries. Their algorithm uses transport layer flow data to identify botnet controllers, and works backwards from detected anomalous traffic. A host with suspicious behaviour is first detected, from which more anomalous flow records are recovered. This is analyzed and aggregated with other flows in the network to find candidate control servers, which can then be further investigated. Because of their reliance on the general structure of botnets rather than any specific implementation, this method of detection applies to many botnets. While it is only applicable to IRC botnets, they believe that similar methods can be made for botnets with other control structures, which includes Mirai.

Metrics designed from the dataset

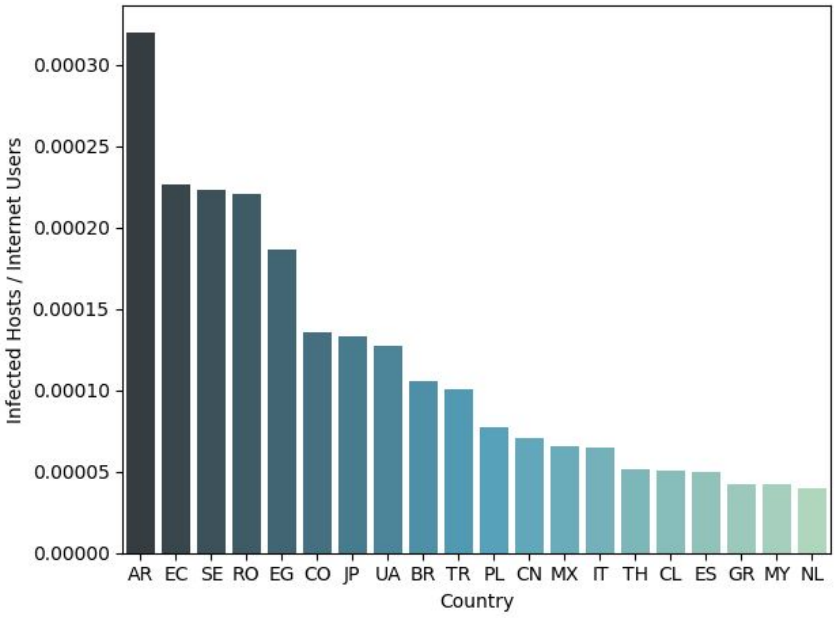
The type of these metrics is *incidents*, because the dataset contains incidents of mirai packets being sent. The data is thus related to the threat environment. Our data only contains information on which IP's are known to be compromised by the botnet, so we cannot derive any metrics related to costs.

We derive three metrics from our dataset, and explain here why we believe they are of use to a security decision maker. In the next section, we compute the results of these metrics and draw conclusions from them.

1. Number of infected IP sightings per country, normalized for users in country
   Why: This metric indicates regions that are relatively more infected and therefore more likely to attempt to infect our system. Knowledge of where the botnet is located is important for prevention and predicting the cost of preventative measures.
2. Number of new infected IP sightings over time
   Why: This metric gives us an idea of how fast Mirai-like botnets are spreading. This can help us in multiple ways. It helps us predict whether the capabilities of the botnet are growing, which means prevention would cost more since more would be needed. It can also help us determine that a new vulnerability is being exploited, as this would lead to a spike of newly infected machines.
3. Amount of infections over country and time
   Why: This metric could show us that the botnet activity focuses on certain regions at a time, indicating possibly targeted attacks or locally spreading infections. This could help predict the risk we currently face by determining whether there are many infected devices near our computing equipment that might attempt to spread there, as well as whether the botnet has a precedent for targeting specific regions/companies.

Evaluating the Metrics

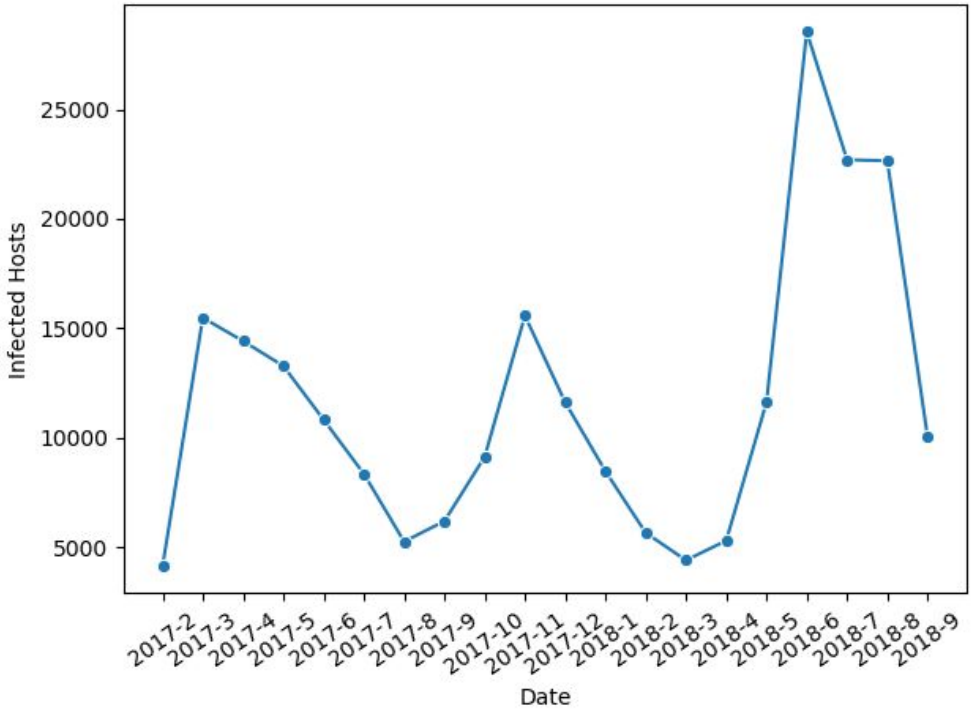
Number of infected IP sightings per country, normalized for machines in country



In this metric we see the total of infected hosts in a country which is then normalized by dividing by the number of internet users in the country (as in 2018) [2]. The countries are selected from the 50 countries with the highest number of internet users. The top 20 countries for this metric are shown. Normalization is important here since without it one could easily confuse high metric results with higher likelihood of infection while the metric could actually say more about the country's number of internet users than about infections.

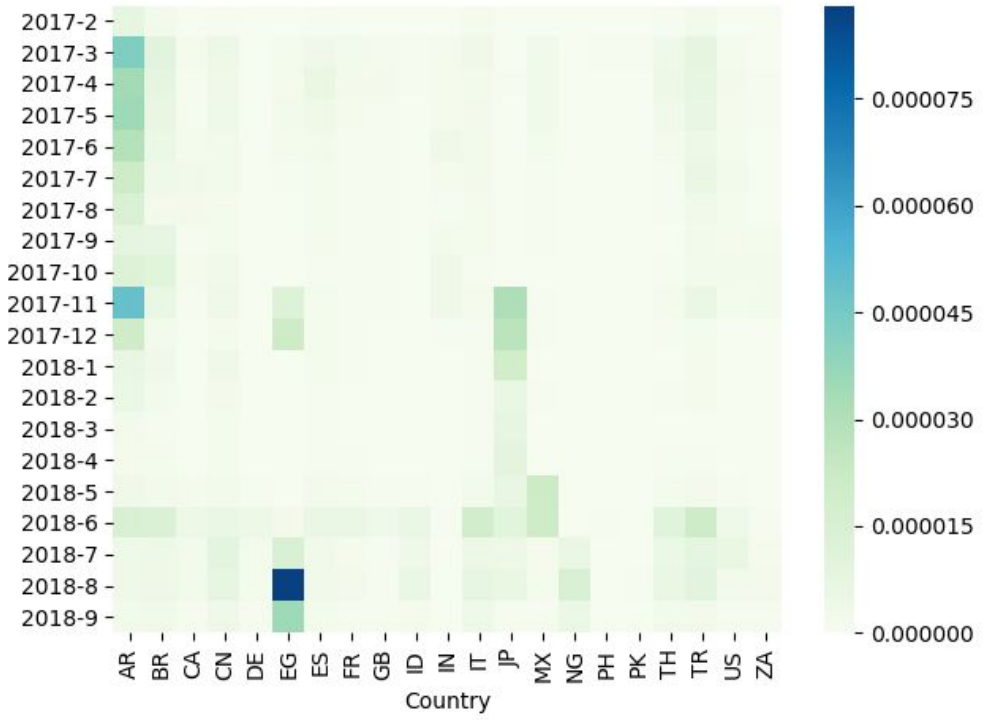
The number of infections per country can be cross referenced with visit data of the services under administration of an IT admin, and could be used to see if there are certain visiting countries that match in rank with the mirai botnet spread to estimate whether bots are actively visiting your services. Furthermore, this graph gives an indication on the security policies in place in certain countries, helping us make more informed decisions when expanding to new computing infrastructure globally. It shows that many south american countries contain a relative high number of bots, which is potentially caused by less strict security policies.

Number of new infected IP sightings over time



In the above figures we plot new Mirai infections per month. The number of infected packets over time can be interpreted as the infection rate of the mirai botnets. What is interesting about this timeseries are the spikes in november 2017 and june 2018, these share some similarities with the spike discussed in size analysis of the mirai botnet in [1]. In that case it was a new exploit which could be used by the mirai botnet to infect more devices, a similar thing could be the case for the peaks in november 2017 and june 2018. Furthermore, this time series is an indicator towards the activity within the botnet and could therefore be of use for IT admins to know if the potential impact from a ddos attack from a Mirai botnet is increasing or decreasing. This can then be used to make intelligent decisions on ddos controls.

Amount of infections over country and time



This metric counts the amount of infected hosts normalized by country's amount of internet users, then grouped by both country and time. This is useful because Mirai infections do not spread equally fast in every country.
By cross-referencing the infections grouped by both time and country with the connections that a service receives a decision maker could get more detailed insights in when they receive traffic from infected hosts. For example in this visualization of the metric we see that in EG (Egypt) the infections occurred later than in AR (Argentina). A decision maker can therefore be more sure that no malicious traffic came from Egypt in early 2017 while they should focus more on Argentina.

References

[1] Bailey, M., Cooke, E., Jahanian, F., Xu, Y., & Karir, M. (2009, March). A survey of botnet technology and defenses. In *2009 Cybersecurity Applications & Technology Conference for Homeland Security* (pp. 299-304). IEEE.

[2] Akiyama, Mitsuaki et al. "A Proposal of Metrics for Botnet Detection Based on Its Cooperative Behavior." *2007 International Symposium on Applications and the Internet Workshops*

[3] Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., ... & Kumar, D. (2017). Understanding the mirai botnet. In *26th {USENIX} Security Symposium ({USENIX} Security 17)* (pp. 1093-1110).

[4] Imperva Security (2016, Oct 26), Technical Analysis of Mirai, Retrieved from https://www.imperva.com/blog/malware-analysis-mirai-ddos-botnet/

[5] Jerkins, J. A. (2017, January). Motivating a market or regulatory solution to IoT insecurity with the Mirai botnet code. In *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 1-5). IEEE.

[6] The CAIDA UCSD, "UCSD Network Telescope" http://www.caida.org/data/realtime/telescope/

[7] Karasaridis, A., Rexroad, B., Hoeflin, D. (2007). Wide-scale botnet detection and characterization. In *Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets*.

[8] Prabhu, T. N. (2019, August 10). List of Countries by number of Internet Users. Retrieved from https://www.kaggle.com/tanuprabhu/list-of-countries-by-number-of-internet-users