WM0824 Economics of Cyber Security
Assignment 1

Maria Simidžioski 4381319

Sam van Hooff 4247620

Tom Slooff 4492919

September 2019

1 Introduction

1.1 The security issue

This study will focus on the security issue of a company which manages and supervises an autonomous system (hosting provider). Their servers are attacked and defaced, and the metrics defined in this study can be used to assess the situation and to measure the level of security of these autonomous systems. Website defacement in this study means the act of changing the appearance of a website or replacing a website with a different one. Thus injecting javascript to mine bitcoins while maintaining the layout of the website is not seen as website defacement. This is distinguished in the dataset.

With a website defacement the integrity of the server is compromised. Important values that are affected are the image and reputation of the company to which the website belongs, as well as the AS company. The trust of consumers in both of these companies can go down, and this leads to indirect costs due to consumers diverting to other companies. The availability of the original website is affected which can lead to direct costs, in the sense of restoring the original website and also because of missing business due to the website being unavailable.
The costs of a defaced website can be estimated by looking at the pricing of protection tools. It is logical that these costs must be lower than the costs of website defacement, meaning that the pricing serves as a minimum bound for the costs. Using the pricing of the sucuri tool, which is 200 - 500 dollars per year depending on the type of plan (Sucuri 2019), the minimum costs start a bit above this price range. However the costs depend largely on the size, image and reputation of the company to which the website belongs.

Website defacement can be an objective of several adversaries. Examples of these are activists who want to share a message, often related to the company which website was defaced, or vandals who want to show what they are capable of.

1.2 Dataset

The given data is collected by internet scanners which scan all sorts of websites across the internet. When a scanner finds a defaced website it collects data about that site. The data includes how long the website was defaced, with a

beginning and an end time, but also data such as the domain, the autonomous system it belongs to, the country it is located in, and more. Some data about the connection is also given, such as the used name servers. This data can be used to create metrics which can be analyzed over time in order to evaluate the security level of the ASs.

2 Metrics

In this section the ideal metrics for security decision makers in organisations that deal with Autonomous Systems will be described and discussed. Also a short overview of different metrics that exist in practice will be given.

2.1 Ideal metrics for security decision makers

The ideal metrics for security decision makers would include all of the four categories of metrics: controls, incidents, vulnerabilities and losses because each of these categories explores a different aspect of the threat environment as stated by (Gañán and Eeten 2019).

In the case of an organisation that supervises and provides Autonomous Systems, the reputation of the organization is critical for its success and therefore it is very important to be able to measure and interpret the level of security of their Autonomous Systems. Since this cannot be directly measured and used to interpret the reputation scores, different metrics are needed that will provide accurate and meaningful results. These metrics need to reflect on the aspects of the level of security and reputation of the organisation.

A metric that could be used is measuring the number of incidents, in this particular case, the number of website defacements that occur in an Autonomous System. However, simply counting the number of incidents and abuses would not be enough. Additional to this, it is very important for a decision maker to also include information about the exposure component. Aspects such as size of the Autonomous System, as well as the number of users, the capability of the adversaries, and other factors that have an influence on the number of incidents.

Another crucial aspect that needs to be included in the metrics is the financial impact and security costs that are related to the incident as proposed by (Gañán, Ciere, and Eeten 2017). They also argue that this can be difficult in practice because different costs exist, such as direct and indirect costs and it is very difficult to generate accurate monetary estimates that will include all aspects.

These issues prevent the possibility of creating an ideal metric that accurately measures the level and costs of security for the organisation.

Lastly the types of adversaries and their motivation is very important in measuring the security to mitigate and prevent these incidents. (Romagna and Hout 2017) argue that the potential adversaries that would be interested in abusing the Autonomous System are mostly hacktivist or individuals with political or religious motives, which often display socio-political messages on the defaced websites.

To summarize, the ideal metric would include measure the number of incidents, the indirect factors that have an influence on these incidents, the financial impact and costs related to the incidents and the type of adversaries that perform these incidents.

2.2 Metrics that exist in practice

Several studies have focused on creating metrics or guidelines for metrics that are suitable for AS supervisors or hosting providers.

(Shue, Kalafut, and Gupta 2012) argue that in practice, most of the organizations that supervise and host these Autonomous Systems use metrics such as a count of the number of viruses/infections in an Autonomous System that is in some cases also normalized by the address space size of the adversary.

(Pieters, Ven, and Probst 2012) have proposed a way to quantify the security risk and elaborated on how to measure the vulnerability level of a system to a specific threat scenario. They measure the three variables related to the vulnerability of a system, namely, the capability of a threat, the strength of a control and the likelihood that a threat is successful for a specific asset. This metric is based on the Elo ratings that are used for ranking chess players (Elo 1978). (Pieters, Ven, and Probst 2012) show that because the threat capability and strength of control variables are explicitly calculated, more accurate prediction can be made about how successful future incidents or threat events will be.

(Noroozian, Korczynski, et al. 2015) discuss two reputation metrics, namely the *Occurrence of abuse* and *Persistence of abuse*. The *Occurrence of abuse* measures the number of abuses in an Autonomous System and a way of achieving this metric in practice is to count second level domain IP pairs instead of just counting the number of distinct IP addresses as the unit for abuse as proposed by (Noroozian, Korczynski, et al. 2015). The other metric *Persistence of abuse*

measures the longevity or the presence duration of the abuse, which gives information on the amount of time it takes for the organization to respond to the abuse.

(Noroozian, Ciere, et al. 2017) also address the challenges of measuring the security performance and develop a model that is based on the Item Response Theory (IRT). A security performance metric is tested and evaluated as a latent variable from a number of heterogeneous datasets. This model takes into account the exposure of the Autonomous System supervisors/providers, one example of such a variable is the size of the Autonomous System.

(Böhme 2010) reviews and discusses the relation between the existing security metrics and the security investment models. He defines the security level as a variable that describes the protection quality. This variable is difficult to measure since it includes both deterministic and stochastic indicators. (Böhme 2010) concludes that the metric that is mostly used in practice by organisations regarding the level of security and the costs of security is the *Return on security investment (ROSI)* which is explained as:

$$ROSI = \frac{\text{benefit of security - cost of security}}{\text{cost of security}} \tag{1}$$

3 Evaluation of the metrics

In this section the created metrics will be described and applied to the given dataset. Lastly the metric will be evaluated.

3.1 Average Lifetime

The average lifetime is a metric to that gives an indication of the awareness and the reaction speed of an Autonomous System (AS) operator. The lifetime is defined as the time difference between the first notice of a website defacement and the time the issue is resolved. Considering the dataset, this entails the difference between 'firsttime' and 'lasttime'. Effectively this gives a picture of the reaction speed of an AS and its ability to swiftly mitigate website defacement issues. And thus reflects the overall health of the net-space governed by the AS operator.

The average lifetime of web defacements of an AS is calculated by obtaining the mean difference between the firsttime and the lasttime over all 'dead' defacements. The notion of a server being 'dead' means the website is no longer

defaced. For this metric to make sense, it is a prerequisite that an AS has enough incident-data to draw conclusions from. If an AS only has had 1 website defacement, it is useless to infer general characteristics of an AS from this sole observation. This is why ASs with less than 100 website defacements have been filtered out of the analysis.

Another issue which needs to be addressed is the occurrence of outliers within the lifetime-data of an AS. There is a risk of outiers driving the average, which would result in scewed conclusions. To reduce this effect, outliers have been filtered out using a z-score. The z-score of an observation illustrates how many standard deviations this observation is from the mean. If the z-score of an observation is above a threshold it is filtered out. A widely used z-score threshold to define an outlier is 3, so this threshold is also chosen in this analysis.

Figure 1 and 2 show the best and worst performing ASs respectively (based on lifetime). These graphs show the average lifetime of a defaced website per AS (blue) and the number of incidents an AS has had (orange). This information is useful in the process of defining best/worst practices. As shown in figure 1 there is only a hand full of ASs capable of resolving defaced website issues on their net-space within a week. There are some interesting players identified in this graph, which could function as best practice. An example of one of this players is AS18978. This AS has an average lifetime of approximately 10 days, while the amount of incidents on its net-space is huge (around 1800). This shows the ability of an AS to respond fast and effectively.

The same principle holds for the 20 worst performing ASs in figure 2. AS16265 and AS6939 have a tremendous amount of incidents, while their ability to mitigate the issues is inadequate; over 2000 website defacement issues with an average lifetime of approximately 200 days.

The average lifetime metric is useful to define the ability of an AS to adequately react to an adversary defacing a website controlled by the AS. This gives a clear picture of the state of affairs of an AS internally, but also of the current landscape and an AS's position in this landscape. This can be used to learn from the actors who are performing better. A sensible next step in this research would be to identify character traits of successful ASs and how this traits affect

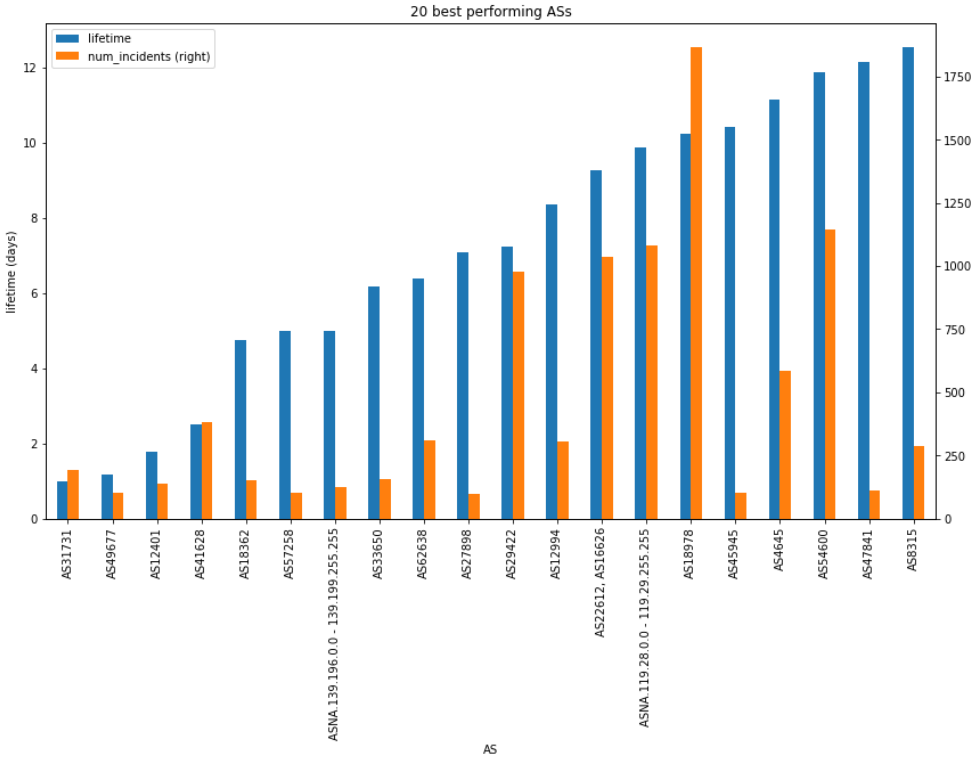


Figure 1: 20 ASs with the lowest lifetime

the average lifetime.

3.2 Normalized Defacement count per Month

The normalized defacement count (NDM) aims to evaluate the situation of the security issue for ASs (autonomous systems) at moments in time. This is done by counting the amount of website defacements for every AS for each month. This is then divided by the size of the AS to account for the size component of the exposure. The size is calculated by taking the amount of IP addresses in the ranges defined by the 'inetnum' field in the dataset.

An intuitive interpretation of the NDM is to take it as the percentage of addresses of the AS that are defaced during the month. However this is not entirely accurate as the NDM does not look at unique websites which are defaced. Attackers often re-deface, meaning they attack the same website twice. This means that the NDM can have values higher than 1.

In order to look at the distribution of values for the NDM, it was averaged for each AS. Two histograms of the average can be seen in figure 3. Figure 3a shows

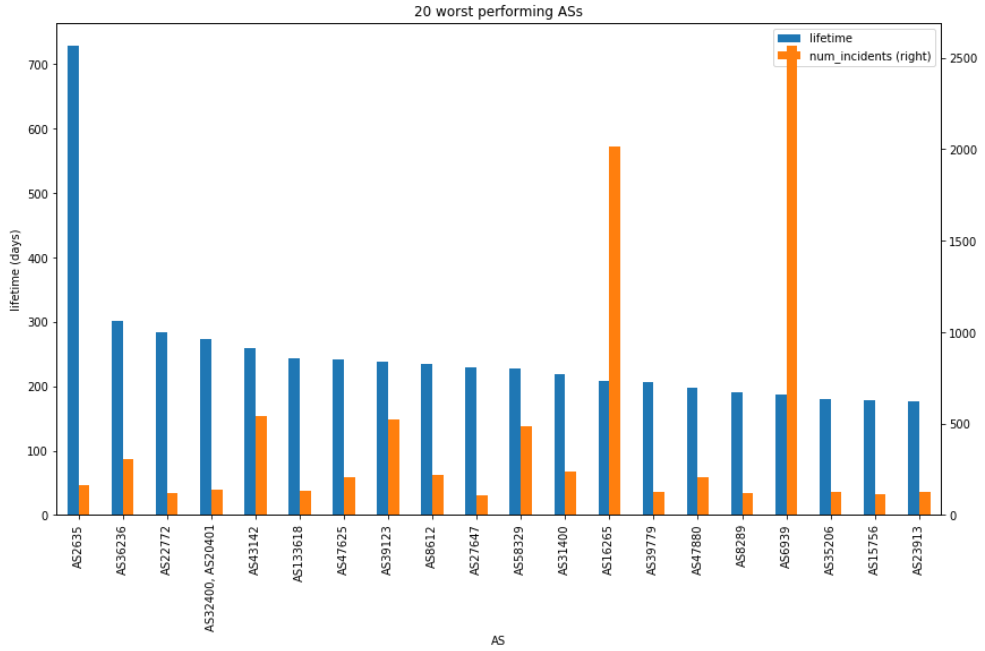
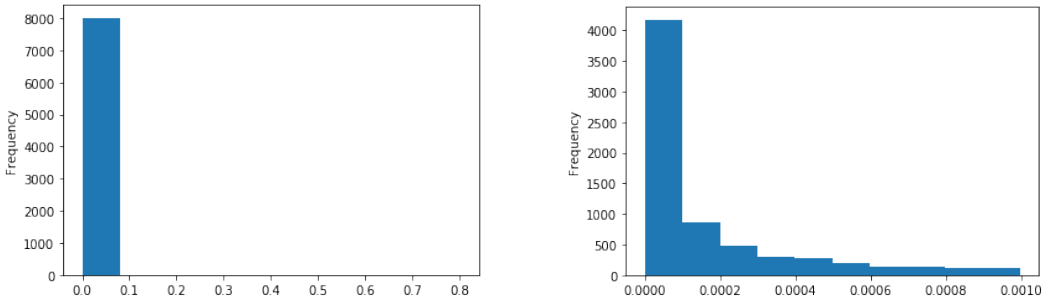


Figure 2: 20 ASs with the greatest lifetime

the distribution across the whole dataset. The values go up to 0.8, but as can be seen the majority of values are below 0.1. Figure 3b shows the distribution for a subset of the data: For values lower than 0.001. Here again the majority of the values are situated on the lower side of the histogram, with most values being below 0.0002. It is clear that the NDM typically has very low values. This can also be seen in the average of the mean, which is 0.00139 (sd = 0.0144).

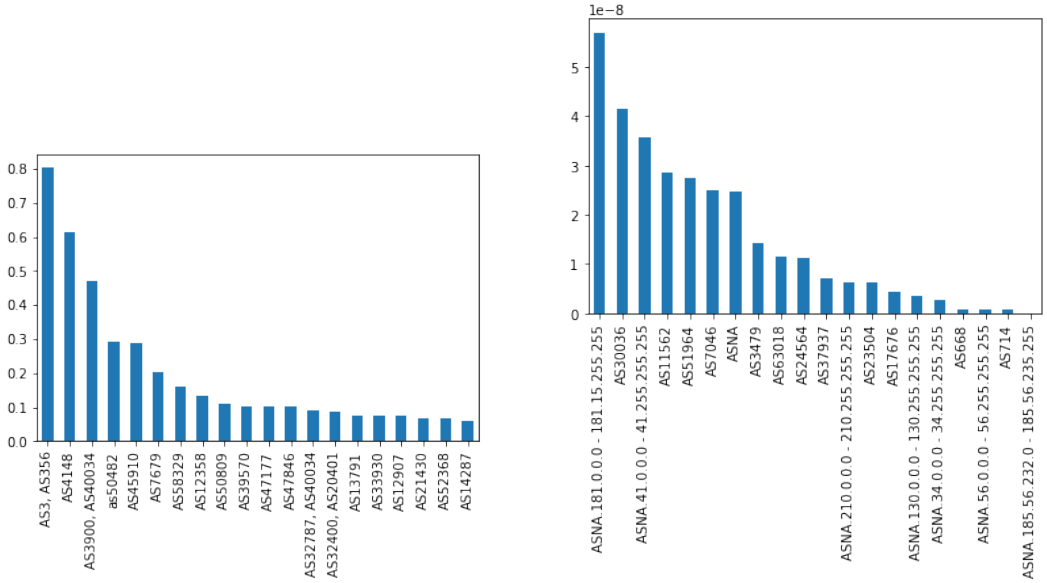


(a) Histogram of complete data

(b) Histogram of metrics up to 0.001

Figure 3: Distribution of the metric

Figure 4 shows the bar plots for the lowest and highest 20 values. Both of these plots have the shape of an exponential distribution, as could also be seen in figure 3.



(a) Bar plot of the highest 20 values

(b) Bar plot of the lowest 20 values

Figure 4: Distribution of the metric

Figure 5 shows how the NDM value changes for the 5 ASs with the highest average metric value over time. This shows how several of these ASs had specific 'outlier' months with an extremely large amount of attacks, with the highest metric value being 10 for AS 4148. The time component is an added benefit of the NDM as this allows comparison across ASs for specific timeframes. The graph shows that all ASs had a spike of attacks between 2012 and 2014, and for all ASs there were then relatively few attacks in the year 2014.

To evaluate the metric an import aspect to note is the nature of website defacement. A natural drawback of using incident count to assess a security situation is that you may not notice every incident. However with website defacement the intention of the attack is to be noticed. Thus it can be assumed that most website defacements will be picked up by scanners.

The metric accounts for the size of the AS, which is necessary for this security issue. The amount of attacks is in itself not a great indicator, if you don't compare it with the amount of servers that the AS is managing. One thing that should be noted however is the calculation of size through the dataset. The size of the AS is calculated using all the ip ranges defined in the dataset. However an ip range of an AS is only added if there is at least 1 attack in that range. Thus an AS may consist of 100 different ip ranges, but if only 1 shows up in this

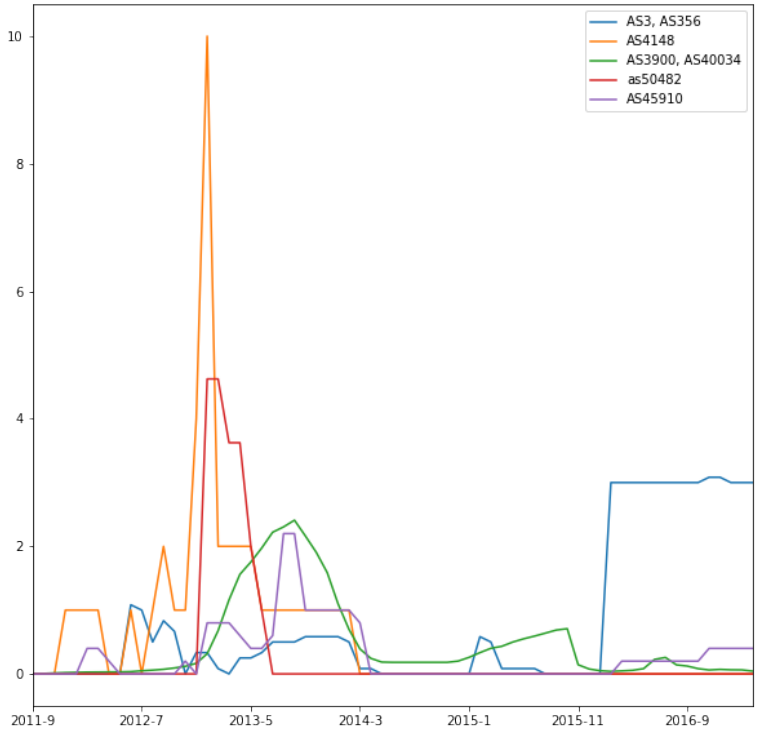


Figure 5: The metric over time for the 5 ASs with highest average metric value

dataset then this is used for size calculation and this AS may seem less secure than it actually is. This is a drawback of the data available, but it nevertheless serves as a useful normalization method.

Lastly the aspect of time in the metric is very useful. As could be seen in figure 5, there seem to be some general trends for all ASs over time. Additionally, keeping the time aspect in the metric makes it possible to look if there were some causes for an increase in attacks for certain periods, e.g. a new vulnerability. For the defined metric a month was chosen as unit because it was the most detailed feasible unit over the time period. The data ranges from 2011 to 2017, so taking a weekly average gives a lot of noise and makes it more difficult to find trends.

As mentioned before, the NDM is based on incidents. This means that some aspects of exposure already affected the value of this metric. Incident count depends on the security of the website, but it also depends on other factors such as the content of the website, whose website it is (which company), and a range of other possible factors such as geographic location. Thus this metric provides useful insight only assuming that other ASs have similar situations, or by comparing specific ASs for which this information is known and similar. If a website

is defaced several times in a month by an extremely skilled hacker after days of work, and a different website is hacked only once but by an unskilled hacker after 5 minutes, based on the incident count the first website may be less secure but this is not necessarily the case. There may be stronger defenses in place in the first website, but it is simply targeted more because of the content of the websites.

To conclude, the NDM is a useful metric which includes several factors making it a useful metric to compare ASs. It can be extended or used with additional data to account for more factors such as geographical location or content of the websites. It is a useful metric especially in addition to metrics based on controls and vulnerabilities.

References

Böhme, Rainer (2010). "Security metrics and security investment models". In: *International Workshop on Security*. Springer, pp. 10–24.

Elo, Arpad E. (1978). *The rating of chessplayers, past and present*. New York: Arco Pub. ISBN: 0668047216 9780668047210.

Gañán, Carlos H, Michael Ciere, and Michel van Eeten (2017). "Beyond the pretty penny: the Economic Impact of Cybercrime". In: *Proceedings of the 2017 New Security Paradigms Workshop*. ACM, pp. 35–45.

Gañán, Carlos H and Michel van Eeten (2019). "Block 2: Measuring Cyber Security". In:

Noroozian, Arman, Michael Ciere, et al. (2017). "Inferring the Security Performance of Providers from Noisy and Heterogenous Abuse Datasets". In: *16th Workshop on the Economics of Information Security.*

Noroozian, Arman, Maciej Korczynski, et al. (2015). "Developing security reputation metrics for hosting providers". In: *8th Workshop on Cyber Security Experimentation and Test ({CSET} 15)*.

Pieters, Wolter, Sanne HG van der Ven, and Christian W Probst (2012). "A move in the security measurement stalemate: Elo-style ratings to quantify vulnerability". In: *Proceedings of the 2012 New Security Paradigms Workshop*. ACM, pp. 1–14.

Romagna, Marco and Niek Jan Hout (2017). "Hacktivism and Website Defacement: Motivations, Capabilities and Potential Threats". In:

Shue, Craig A, Andrew J Kalafut, and Minaxi Gupta (2012). "Abnormally malicious autonomous systems and their internet connectivity". In: *IEEE/ACM Transactions on Networking (TON)* 20.1, pp. 220–230.

Sucuri (2019). URL: https://sucuri.net/website-security-platform/signup/.