Economics of Cyber Security:
Botnet Tracker: Factors influencing security metric
Assignment Block 4

M. Kesteloo - 4291158
R.D. Meeuwissen - 4287150
P. Remeijsen - 4286243
B. Czaszyński - 4571894

October 22, 2018

1

INTRODUCTION

In this report we analyze the security metric defined in the previous assignments and determine which factors influence the variance in the metric for the different actors. The focus will be mainly on identifying what differences in actor environment and potential risk strategies impact costs and benefits of certain countermeasures.

The case study for this research is botnet networks among providers of Internet infrastructure. This encompasses companies delivering connectivity and providing domain-related services.

Botnet infections and seizure of control over a large group of computers, will be the security issue analyzed in this research. Dependent on the actor impacted by its influence, distinct countermeasures with varying levels of influence and mitigation are presented.

2

ACTORS INVOLVED

The following section presents actors whose operations are significantly impacted by the existence and spreading infections of botnet networks. Each of the actors provide different functionality to the environment of online businesses and suffers different risks due to threats posed by botnet networks. We have identified the hosting provider as the problem owner and proposed two additional entities, due to their significant role in the botnet ecosystem.

2.1. INTERNET SERVICE PROVIDERS

ISPs are the entry gates to the Internet. Each bit of traffic is transferred through their systems, which opens possibilities of taking action against the security issue at hand. Botnet nodes obtain commands from a central server responsible for issuing and delegating tasks to the botnet participants. Capability to inspect all the traffic from every machine connected to the Internet in a certain world region allows to easily classify communication patterns, estimate botnet sizes and re-actively engage in counter-activity.

In order to analyze data in transfer, the ISP has to engage in performing deep packet inspection. It is a method of investigating data contents of every packet sent over the underlying network infrastructure. As most of the current Internet traffic is sent in encrypted form, the access point of the service provider would have to act as a proxy SSL server and, through means of providing an intermediary certificate, effectively perform a man-in-the-middle attack on every connection. This grants the possibility of performing traffic classification, identifying malicious activity and employing a plethora of methods mitigating the risk posed by active botnet machines.

Enabling such a deep insight into type, purpose and content of online communication comes with a set of serious drawbacks. Utilization of deep packet inspection methods together with complete unfolding of SSL encryption causes a serious privacy issue. Before intrusive traffic classification happens, the ISP has no means of accurately determining which portion of the traffic it routes comes from legitimate users and which is a result of malicious botnet activity. The access point provider effectively intercepts and reads all communication of all its clients. Implementation of such investigation techniques, although potentially profitable from data-oriented business opportunities, or espionage activities, would cause an enormous social opposition and irreparable reputation damages.

ISPs are the least incentivised among listed actors to act upon botnet threat and implement the aforementioned countermeasure. Potential loss of customers and negative reputation backlash cause a bigger threat to their existence, than risks associated with botnet infections and their further operations. ISPs are not a frequent target for ransomware attacks either, as they are the backbone of Internet infrastructure and obstruction of their correct operations would significantly limit potential reach of bots.

The external actors interested in mass data collection and profiling users on big scale would be the ones most interested in utilization of such security techniques by ISPs. An incident of this type has been uncovered by Snowden's revelations from 2013, where governments and technological giants have been exposed

for monitoring traffic at the biggest Internet exchange points on the planet. Without need for much infrastructure adjustment and virtually no way to reveal such activity, besides insider threat, espionage and mass-monitoring could be enabled on a world-wide scale. However, using such techniques would be devastating for cyber criminals using botnets because with deep packet inspection, malicious code will be intercepted significantly more often than without deep packet inspection. This way, the criminal will not be able to infect many devices and thus will end up with a small botnet. When ISPs do not take countermeasures, botnets will still be a negative externality. Criminals can for example carry out DDoS attacks using the infrastructure of the ISP.

2.2. HOSTING PROVIDERS

Hosting providers have been identified as the main problem owner and an actor with the greatest level of potential influence in solving or remedying the issue of botnet infections. This actor provides physical infrastructure, including computer machines and networking solutions, as well as software and administrative services.

Botnet infection undermines quality of service provided to host's clients. If the host cannot ensure stability of the rented infrastructure, or guarantee expected functioning of hardware and software, many of its clients will turn away and invest in more secure competitors. Such consequences can be avoided by implementing security measures which will increase resilience of the system and lower the chance of infections, effectively providing more reliable solutions to clients. Hence, hosting providers have a strong incentive to invest in security solutions and actively fight threats associated with botnet infections.

Some of the mitigation strategies applicable to hosting environment include hardening of machine configurations on which the servers are running, or employing various anti-malware solutions and intrusion detection systems. One of the most effective countermeasures would encompass a hybrid system, composed of a monitoring module and a reactive module. In this setting, hosting providers could detect and identify malicious behaviour and re-actively carry out steps outlined in their incident response plan. Such a solution can additionally provide shareable information, which when exchanged with other entities having the same incentive in fighting botnets, can aid the cybersecurity community via open-source intelligence repositories. With enough actors contributing to such cause, overall costs of protective measures can be reduced and novel methods and techniques of protecting critical business resources will be developed and improved. Furthermore, closer cooperation and tighter bounds between companies working in the same branch of market will decrease the quality gap and promotes healthy and balanced competition.

Incorporation of any mentioned countermeasure would require significant financial investment. AV software would have to be installed on every running machine which is directly tied to purchasing a bulk of commercial licences from the vendor, increasing vendor revenue. Reconfiguration of systems and proper hardening requires employment of a specialist IT operator. Effects of such investments would be a considerable increase of system's resilience to any type of attack, not necessarily focused on prevention of botnet infections. Less unwanted network traffic would also allow the provider to allocate more network bandwidth to clients, increasing capability of servers without investment in hardware upgrades. Regarding botnets, C&C servers are hosted on the hosting provider's machines which may cause other websites hosted on the same machine to become infected as well meaning this is a negative externality. If the hosting provider chooses to remove the C&C servers from their machines, then the infected devices will be useless to the criminal as the C&C server will not be able to control those devices anymore.

2.3. ANTI-VIRUS COMPANIES

Every company conducting their business through an online platform or providing any type of digital services on the Internet has to employ certain levels of security. Correct functioning of an underlying infrastructure is critical to their business operations and continuity. As the main focus is usually on what the company specializes in, little effort can be put on handling cybersecurity on their own. Hence, security solutions are either bought from another vendor or completely outsourced. This introduces reliance on solutions of an external actor, whose product's performance impacts or completely determines if a business can operate in a desired way.

Anti-virus (AV) software development companies act as a vendor taking over responsibility of ensuring correct and uninterrupted operations of their clients' systems. There are many AV software development companies which sell their AV software to clients, so clients can choose from a wide range of available software. To get a piece of the cake, AV companies should have a malware detection rate similar or better than their competitors. If their detection rate is lacking in comparison to others, they will lose customers and hence lose money. The main incentive for improving security performance is thus financially motivated.

The countermeasure which directly affects the security metric of choice is improvement of malware detection rates in a vendor's product. As this is the main selling commodity of the AV companies they have a strong incentive to constantly improve their product, not only to provide a reliable solution to the client, but also to ensure a competing position on the market. Since this falls within standard business practices of such vendors, no specific business operation adjustments have to be made, besides additional funding. Increased focus on research and development and potential expansion of hired staff is all the vendor company shall do in order to employ the discussed countermeasure.

2.4. EXTERNALITIES

Regarding externalities around the security issue, it is hard to pinpoint what exactly might the outcomes be. Given appropriate technological advancements are made by the research and development department, the whole computing industry can become more secure, due to increased detection rate and necessity to put more effort into breaching security perimeters of clients utilizing the AV solution. On the other hand, more effort means that only targeted attacks and advanced persistent threats might become profitable to adversaries, which might result in an increased amount of breaching attempts on critical infrastructure. Higher success rate for detection of old malware can result in more research efforts on the adversaries' side, resulting in discovery of new intrusion techniques, or even higher level of professionalization of black-hat hacking.

One of the most significantly impacted externalities is the botnet itself and the environment in which it operates. As the countermeasures are implemented to directly neutralize threats, or partially mitigate risk existing due to botnet operations the externality is categorized as a negative type. Implementation of different aforementioned techniques can cause varying levels of disruption in normal operations of a botnet. Disconnection from the rest of the infrastructure and most importantly command and control server can render affected botnet nodes no longer active and incapable of malicious behaviour.

3

Security Performance

3.1. Metrics

The metric we will use is the percentage of infected domains per hosting provider (see Figure 3.1). The names of the hosting providers are found by mapping the ASN numbers from the ransomware dataset to the hosting provider. The metric shows that *FC2 INC*, *Confluence Networks Inc*, *CyrusOne LLC* and *Namecheap, Inc* have the largest infection rate. In the next sections it will be investigated what factors could cause this.

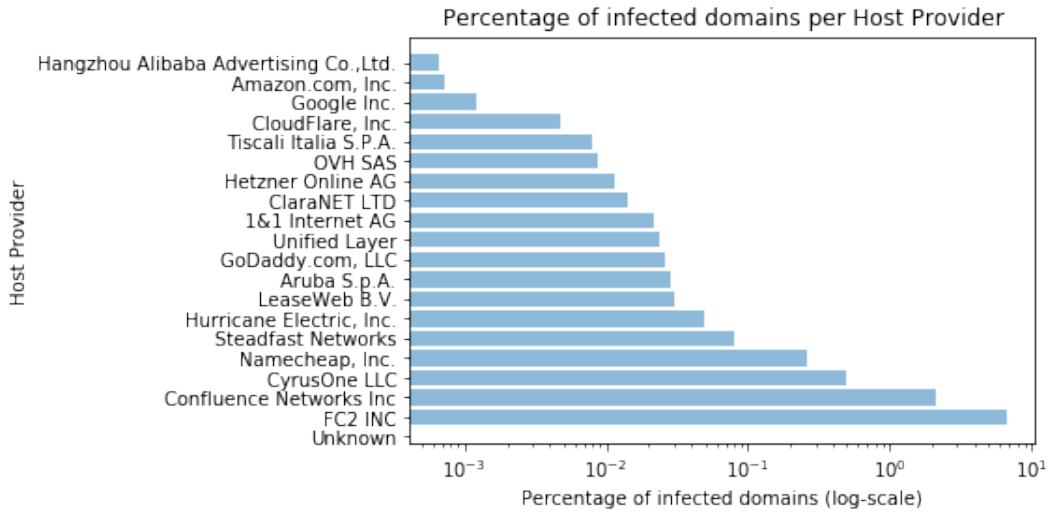

Figure 3.1: Graph of the percentage of infected domains per hosting provider

3.2. Factors Causing Variance

After finding variance in the results for the host provider, we can start to discuss the possible explanations of the difference in performance. We can differentiate between factors that are hard to quantify, leading to problems when attempting to find correlations between the factor and the metric. A number of these factors could be of influence such as:

- **Awareness training among employees**: This might seem like a simple solution. However, simply making the employees more aware of the dangers that might present themselves will help to minimize the number of incidents. An example could be to instruct employees not to open attachments on e-mails form an unknown source.
- **Firewalls and other types of monitoring:** This is an example of a factor that differs from company to company and has a great influence on the security. The fact that different companies choose different

vendors could be explained by a number of factors such as available budget and organizational priorities. Furthermore, some hosting providers might not have the manpower to actively handle monitoring activities. This can lead to longer times between infection and reaction.

- **Anti-virus software**: Similar to the previous point there can be variation due to the budget and organizational priorities. The running of the software might affect the operational infrastructure, so it will be a trade-off between security and

However, these factors are hard measure and few information is available about these factors. One of the reasons for this is information asymmetry. Security companies do not want to disclose information about negative effects of their products. Retrieving this information for public research is therefore hard. Therefore for this assignment, it makes more sense to look at factors that can be quantified. Research by Tajalizadehkhoob et al. shows a couple of these factors [1]. Listed below are quantifiable factors which we have chosen to include in our analysis and determine relationships between them and the security metric investigated.

- **The rule of law** per country is retrieved from the current rule of law index values from the World Justice Project (WJP)[2]. The WJP is an independent, multidisciplinary organization working to advance the rule of law worldwide. They explain their selves in the following way:

  > *Effective rule of law reduces corruption, combats poverty and disease, and protects people from injustices large and small. It is the foundation for communities of equity, opportunity, and peace—underpinning development, accountable government, and respect for fundamental rights.*

- **E-Participation**: E-Participation is described by Macintosh[3] in the following way:

  > *The use of information and communication technologies to broaden and deepen political participation by enabling citizens to connect with one another and with their elected representatives.*

  It describes the the amount of ICT related processes involved in government and governance. This factor could give some insight whether ICT processes involvement in the government helps reduce botnets inside a country.

- **Extent of staff training**: The extent of staff training is defined as an extent to which companies invest in training and employee development. It is related to the awareness training factor described earlier. However, in this case it is seen as an investment on country level.

- **Capacity for innovation**: The capacity of innovation is defined as an organizational potential to innovate [4]. Instead of an organizational potential we focus on the potential of innovation a country level. A country capable of innovating, can direct more resources to developments and improvements, as well as introduction of additional technological solutions boosting reliability of provided services. High capacity for innovation indicate ability to grow and shift focus from simply providing the basic service to enhancing the quality and reliability of the service.

- **Amount of internet users**: The amount of internet users depicts a percentage of individuals having access and actively participating in Internet-based services. Higher involvement indicates higher levels of infrastructure integration and more dependence between various service providers. This results in a more complex networking and potential of more impactful compromise.

- **Secure internet servers**: The security of internet servers is a measure of Internet servers using encryption, per million population. This indicator can play a major role in variance of botnet infection numbers. If all internet servers would be secure, that would mean it harder to be compromised by malware, but also analysis of the traffic and employment of some countermeasures can be less effective.

3.3. DATA COLLECTION

As discussed in the previous section, collecting data from some factors, which would give interesting insight, is hard because companies are not willing to make this information public.

We can however get a lot of factors on country level. A lot of countries are willing to cooperate with each other and make a lot of aggregated statistical data public. In the rest of this section it is discussed how data for these factors was collected.

The rule of law index can be retrieved from the World Justice Project website[1]. Adding this Rule of Law index to the ransomware dataset we can get more insights in this factor of variance. In Table 5.2 all host providers, found in the metric about the infection-rate of the top 20 host providers (see Figure 3.1), are divided by the country where they host and the amount of infections of that host provider in the corresponding country. From all those countries the sum of infections of the host providers shown in the metric is shown in Table 5.1. In the last column of both tables the Rule of Law (RoL) of the corresponding country can be seen. In the next section it is analyzed what the correlation between the infections and RoL is.

Data on factors of E-Participation, extent of staff training, capacity of innovation, amount of internet users, and number of secure internet servers have been obtained from [5]. After aggregation of appropriate information from the ransomware dataset into datapoints per hosting provider, we have grouped them together by country codes to enable integration with data taken from the aforementioned report. The resulting dataset presents number of infections from the same hosting provider and corresponding security factors per country.

3.4. STATISTICAL ANALYSIS

With the data collected, a statistical analysis was performed. For multiple factors defined in Section 3.2 the Pearson correlation coefficient is calculated between the factor and the amount of infections of the countries where the host providers from our metric are based. With this correlation coefficient, one usually determines in advance what the p-value should be. The p-value indicates the probability of concluding that there is a correlation but in fact, there is no correlation. This gives some indication of confidence in statements based on the correlation coefficient. A p-value of 0.05 has been chosen for this study. When the p-value of the calculation is below the selected threshold, the result is statistically significant. The correlations can be seen in Table 3.1.

The table shows that the capacity of innovation has the highest correlation coefficient. It also has p-value below 0.05 which means it is statistically significant. All other factors have a quite low correlation compared to the capacity of innovation factor. Also the p-values are high, which means they are less statistically significant.

Uncertainty
The dataset used for the calculation of these coefficients is rather small which is one of the reasons why the p-value is so high for most of the factors. We could have included all the data from the ransomware dataset, but the data from the factors was not available in a ready-to-use format. We had to retrieve the data by looking it up in the report of the World Economic Forum [5].

In order to show that increasing the number of data points will improve the p-value, we have chosen to include all of the data points for one of the factors, namely the rule of law. This factor was easier to include fully than the factors from the World Economic Forum report. Using all the ransomware data, we get a correlation coefficient of 0.209 with a p-value of 0.092 meaning that there is a small correlation but it is not statistically significant according to our p-value boundary of 5%. However, this p-value is already a lot closer to 0.05 meaning it has more statistical significance than the correlation coefficient reported in Table 3.1.

When taking as many data points as with the other factors, we see that the Pearson correlation coefficient is only 0.098 with a p-value of 0.7069 meaning this result is not statistically significant. Although, we still do not have statistical significance, we demonstrate with this example that adding more data points will improve the results. When calculating the Pearson correlation coefficient, it is important to have as much data as possible.

[1] See https://worldjusticeproject.org/our-work/wjp-rule-law-index/wjp-rule-law-index-2017%E2%80%932018

# infections	**RoL**	**E-Participation**	**Staff training**	**Cap. of innovation**	**% of internet users**	**Secure servers**
Pearson Corr	0.098	0.2895	0.2210	0.5218	0.2151	0.2477
p-value	0.7069	0.2597	0.3939	0.0329	0.407	0.3377

Table 3.1: Table presenting Pearson correlation and p-value results for each correlated pair of factors impacting security metric.

4

CONCLUSION

Based on the results of statistical analysis performed in the study we can conclude that many of the investigated factors do not have a direct relationship to the security metric researched. Most of the factors exhibit very low levels of correlation with high probability of wrongfully rejecting the hypothesis of independence (values of p-value indicate unsatisfactory levels of statistical significance).

The only factor which has been determined to have a strong enough relationship with used security metric is country's capacity for innovation. This factor serves as a measure of the extent to which country's infrastructure and technological development is mature enough to open new opportunities for advancement and investments in novel improvement strategies. Countries reach high levels of innovation capacity upon having access to educated and trained labour force, high availability of resources for local businesses and entrepreneurs, positive local economic growth indicators and high levels of economic well-being.

High dependence between amounts of infections and innovative capacity point towards late prioritization of security goals in organizations. Only upon reaching considerable amounts of spare resources and implementing primary functionality of offered services, a higher priority and focus is given to enhancement of cybersecurity solutions in organizations. Data provided in [5] indicate that ensuring security of infrastructure requires capabilities of affording innovative technologies, more than relying on other factors. As the ecosystem of botnet malware is highly dynamic and adversaries have to engage in a constant chase for newer and previously unseen threats and exploits, the defensive site needs to innovate in order to catch up to the ever changing threat environment.

The investigation of the security metric and factors potentially affecting it revealed a relation with innovative focus and capabilities of organizations on country level. This points toward a need to incorporate novel security technologies in any type of Internet-based service in early phases of the development to facilitate secure operations and resilience against network-based threats.

4.1. FUTURE WORK

Due to very restricted dataset used in the research, many results obtained are statistically insignificant. This drawback could be relieved by performing the analysis on a more coherent dataset with higher focus on hosting providers than on counties where the businesses are operating. This would allow to draw more direct and reliable conclusions on the ecosystem of the problem owner, instead of relying on third party data drawn from a different population sample.

5

APPENDIX

5.1. TABLES

	CountryCode	#infections	RoL
0	AR	2	0.58
1	AU	1	0.81
2	AU	1	0.81
3	BR	1	0.54
4	CA	20	0.81
5	CN	79	0.50
6	DE	219	0.83
7	ES	25	0.70
8	FI	1	0.87
9	FR	165	0.74
10	GB	102	0.81
11	IT	241	0.65
12	NL	116	0.85
13	PL	13	0.67
14	PT	65	0.72
15	RU	5	0.47
16	SG	15	0.80
17	TR	1	0.42
18	US	1584	0.73

Table 5.1: Table showing the amount of infections and Rule of Law index of each country found in the top 20 of infected hosting providers.

	HostProvider	#infections	CountryCode	RoL
0	CloudFlare, Inc.	6	US	0.73
1	Amazon.com, Inc.	17	US	0.73
2	Google Inc.	79	US	0.73
3	OVH SAS	2	AR	0.58
5	OVH SAS	1	AU	0.81
6	OVH SAS	1	BR	0.54
7	OVH SAS	20	CA	0.81
8	OVH SAS	8	ES	0.70
9	OVH SAS	1	FI	0.87
10	OVH SAS	165	FR	0.74
11	OVH SAS	12	GB	0.81
12	OVH SAS	1	IT	0.65
13	OVH SAS	3	NL	0.85
14	OVH SAS	13	PL	0.67
15	OVH SAS	13	US	0.73
16	CyrusOne LLC	110	US	0.73
17	Namecheap, Inc.	1	GB	0.81
18	Namecheap, Inc.	97	US	0.73
19	Hetzner Online AG	136	DE	0.83
20	Hetzner Online AG	5	RU	0.47
21	GoDaddy.com, LLC	15	SG	0.80
22	GoDaddy.com, LLC	470	US	0.73
23	Aruba S.p.A.	85	IT	0.65
24	Steadfast Networks	131	US	0.73
25	Hangzhou Alibaba Advertising Co.,Ltd.	79	CN	0.50
26	Confluence Networks Inc	24	CH	NaN
27	Confluence Networks Inc	56	US	0.73
28	Confluence Networks Inc	153	VG	NaN
29	FC2 INC	138	US	0.73
30	Unified Layer	156	US	0.73
31	LeaseWeb B.V.	1	GB	0.81
32	LeaseWeb B.V.	113	NL	0.85
33	LeaseWeb B.V.	1	TR	0.42
34	Hurricane Electric, Inc.	293	US	0.73
35	ClaraNET LTD	1	ES	0.70
36	ClaraNET LTD	86	GB	0.81
37	ClaraNET LTD	65	PT	0.72
38	1&1 Internet AG	83	DE	0.83
39	1&1 Internet AG	16	ES	0.70
40	1&1 Internet AG	2	GB	0.81
41	1&1 Internet AG	18	US	0.73
42	Tiscali Italia S.P.A.	155	IT	0.65

Table 5.2: Table showing the amount of infections of each hosting provider (found in the top 20 of infected hosting providers) per country with the corresponding Rule of Law index of that country. The "NaN" values are the countries of which we have no Rule of Law information.

BIBLIOGRAPHY

[1] S. Tajalizadehkhoob, C. Gañán, A. Noroozian, and M. v. Eeten, *The role of hosting providers in fighting command and control infrastructure of financial malware,* in *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security* (ACM, 2017) pp. 575–586.

[2] W. J. Project, *Advancing the rule of law worldwide,* https://worldjusticeproject.org/, accessed: 19-10-2018.

[3] A. Macintosh, *Characterizing e-participation in policy-making,* in *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on* (IEEE, 2004) pp. 10–pp.

[4] D. I. Prajogo and P. K. Ahmed, *Relationships between innovation stimulus, innovation capacity, and innovation performance,* R&D Management **36**, 499 (2006).

[5] World Economic Forum, *The Global Information Technology Report 2015 - ICTs for Inclusive Growth,* (2015).