Spamhaus

Externalities

Felix W. Dekker, Nick Ho Sam Sooi, Jorai Rijsdijk, Jakob S. Kok

Group number 6
GitHub https://github.com/Erackron/EoC/

1 Introduction

In previous reports, we have suggested a metric that shows the infection rate per botnet per country. While the metric is normalised in terms of the countries' sizes, it has not been normalised with respect to the botnets' sizes. Therefore, the metric is only really useful when looking at which countries are disproportionately affected by some botnet. In Figure 1 we show some measurements using this metric and observe that Vietnam has a very high rate of Gamut infections. Because of this, we have focussed on Vietnam's Gamut problem.

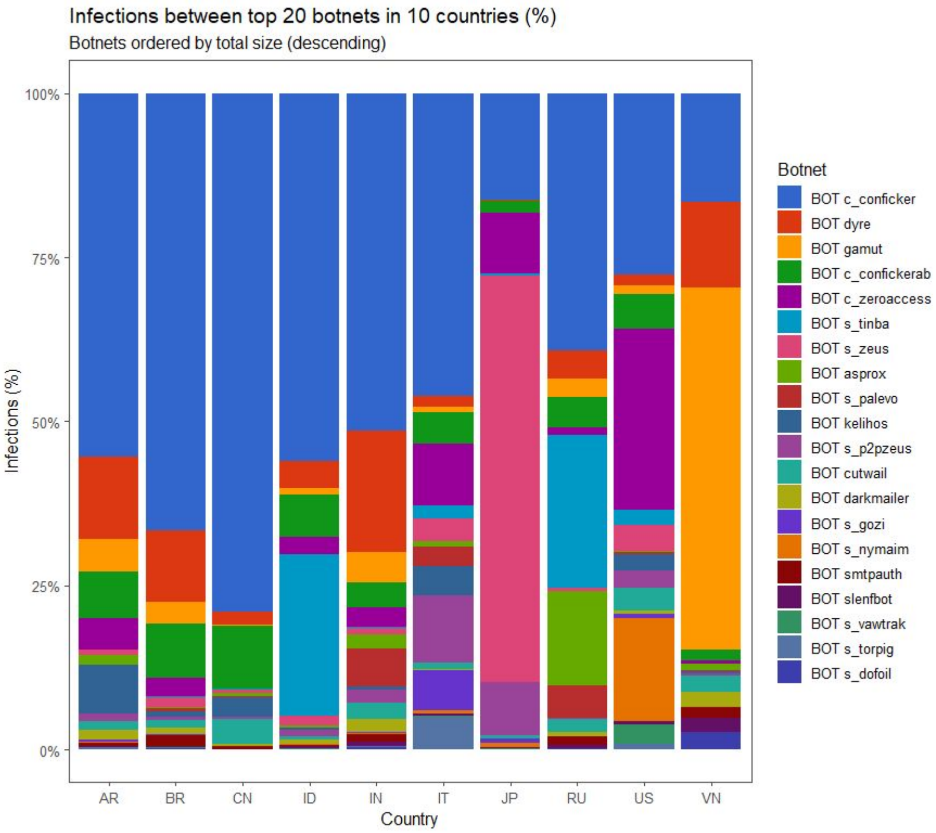

Figure 1. Relative infection rates per botnet per country for the 10 most-infected countries and the 20 most-infecting botnets.

In this report, the previously-described Spamhaus dataset will be used to identify which externalities can be expected when mitigating the risk imposed by the Gamut botnet. Externalities are costs or benefits that affect an external party which does not participate directly in the transaction (Dahlman, 1979). For example, when a production company chooses to dump its waste in a nearby river, the health issues this causes for citizens living downstream can be seen as a (negative) externality. These externalities can cause a lack of incentives for actors to take action and mitigate harm they are causing. In the case of the production company, the company does not bear the costs of the health issues and therefore has no incentive to mitigate risk it imposes for citizens living downstream.

In our dataset, the analysis of the security performance for different countries has shown a significant variance in the infection rate among the countries. In Section 2, we will take a deeper look at the actors involved in this issue and at the externalities their actions may cause. Section 3 will be aimed at defining factors that cause this variance and we will provide a statistical analysis showing the variance that be can be explained by a subset of these factors.

2 Externalities

In this section, the externalities that occur when certain actors invest in the treatment of the risk caused by the Gamut botnet will be assessed. First, three focal actors are chosen. Then, a countermeasure is discussed for each of these actors. Moreover, the distribution of costs and benefits for each actor and the respective countermeasure is evaluated. Finally, the lack of incentives to take the countermeasures as well as the externalities that occur are presented.

2.1 Actors

The first actor we will focus on is the Vietnamese governmental **cybersecurity agency**, which was also the problem owner in our previous report. This agency is responsible for assessing the country's cybersecurity and it reports its findings to and gives advice to the Vietnamese government. The government may, in turn, impose certain security standards on the country's companies and may create regulations and legislation.
The choice for Vietnam is based on the country's disproportionate amount of infections by the Gamut malware. This makes it an interesting point of focus when it comes to analyzing the underlying reasons behind the existence of this problem.

A Vietnamese **IT company** such as VNPT[1] provides a large variety of services, ranging from telecommunications to consultancy. The Gamut botnet poses a threat to VNPT by sending extremely large volumes of spam, which may clog the network and reduce the available bandwidth for legitimate users. If users are significantly affected they may blame VNPT and switch to another provider. Therefore, VNPT should be interested in countermeasures even though Gamut does not directly attack the company.

[1] http://vnpt.vn/en

The machines of **citizens** are interesting targets for malware such as Gamut because these machines are often poorly protected—they use default passwords and vulnerabilities are not often patched. On the other hand, citizens are not directly affected by malware such as Gamut other than slightly reduced performance. Therefore, citizens are an interesting actor to analyse because they can potentially solve the problem but do not have a vested interest in doing so.

2.2 Countermeasures

As the **cybersecurity agency** in Vietnam has general incentives for amplifying the defences of the country's network, one countermeasure that the cybersecurity agency could take to mitigate the security issue is to set up a collaboration network for ISPs. In general, most companies have not been adept at collaborating to defend against cyber attacks, and cybercriminals have exploited the fact that the defenders have not been coordinated. By exchanging information through this new information channel set up by the cybersecurity agency, the companies can to a larger extent coordinate responses and share experiences. This will lead to a more efficient security performance.

The collaboration network could be set up in different ways. Firstly, it could be implemented as a "bond" of companies that allows them to exchange information. Access to the network requires the company to be registered and approved by the government. In addition to this, smaller companies are not allowed to the network meetings because they are not relevant, however, reports based on the meetings are published regularly and allow for the companies to gain an understanding of new developments.

The network could also be implemented as a database where registered companies can list attacks, therefore allowing other companies to gain insight into the current aggregate attack surface. Since most companies will be attacked, a lack of contribution should result in a loss of access to the list. Furthermore, such a system should be supported by a non-disclosure agreement, in order to prevent companies from using the information to gain a competitive advantage. This issue could also be solved by only showing aggregate information to the companies.

IT companies could set up spam filters. These will prevent spam emails from entering the network, thus protecting the network from further infections. Spam filters generally have very high precision, and as we can recall from the analysis in our previous report, it is a cost-effective method against spam.

In order to respond to the botnet threat, **citizens** could properly secure the machines they own. One specific countermeasure they could take against the botnet threat is installing antivirus software. In some cases, the antivirus software will be able to detect and remove the malware from a device. However, in other cases, the complexity of the malware and creativity of the attacker disables the antivirus software from solving the issue.[2] To illustrate

[2] http://clico.pl/services/practical-defense-in-depth-protection-against-botnets

this, many malicious programs present on infected machines are able to interfere with the antivirus software and disable it (Hachem, Mustapha, Granadillo & Debar, 2011). In these cases, prevention in the form of e.g. better passwords may be a better alternative.

2.3 Cost and benefit distribution

In this section, the distribution of costs and benefits for each countermeasure will be discussed.

First of all, the **cybersecurity agency** encounters the sunk costs of setting up the collaboration network. The height of these costs is highly dependent on the way in which the network will be set up. Moreover, the agency experiences recurrent costs related to the facilitation of the network, for example, costs for hosting the meetings or hiring HR staff to coordinate these meetings. The benefits of the network consist of a reduction of the number of infections due to the sharing of defence strategies, incident data and other experiences.

Setting up spam filters in **IT companies** requires an initial investment to research which spam filters have the highest performance levels and hire personnel to install these filters. When making use of proprietary spam filters, companies also experience recurrent costs in the form of a monthly or yearly fee. These investments can be seen as sunk costs since they cannot be recovered. Similar to the collaboration network, the benefits of spam filters involve a reduction in the number of infected machines. Companies which do not invest in the spam filters might also benefit from the investment since these spam filters also indirectly protect other networks from being infected. This effect will be discussed in more detail in paragraph five.

For **citizens**, anti-virus software can either be free of charge or cause a monthly recurring cost for citizens. Although a free countermeasure is attractive for citizens, they often lack features besides simply scanning for known file threats, a technique that most malware actively circumvents, thus limiting the effectiveness of free anti-virus solutions. Paid anti-virus licenses are recurring, sunk costs since citizens cannot sell the licenses and most licenses involve a monthly or yearly fee.
While setting up a better password may take only a few minutes, users with little technological experience may require a few hours and support.

2.4 Incentives

The **cybersecurity agency** aims to improve the security and defences in the country on a general basis. Therefore, an incentive exists for the agency in Vietnam to prevent lost efficiency and productivity in the country due to domains being on blocklists. The cybersecurity agency might also be concerned with the reputation of its national servers.

When dealing with a lot of incoming spam without having a spam filter, the employees of **IT companies** will take much longer to go through their email on a day-to-day basis, leading to

decreased efficiency and productivity and the potential of emails being missed in the flood of spam.
The other side of the incentive to install a spam filter is that if your network is perceived to be sending a lot of spam, your network might end up on a blocklist, which means that other companies with spam filters will likely mark all your legitimate email as spam as well, leading to decreased revenue and a loss of clients.

When a **citizen**'s home network or machine is infected with a botnet, the botnet will take up processing power, network bandwidth, and more. In addition, the botnet may install other malware on the machine, further increasing the effect. Finally, the malware may steal payment credentials, personal information or prevent access to personal data unless a fee is paid. These negative effects might incentivise citizens to properly protect their machines against botnets.

2.5 Externalities

The cooperation network set up by the **cybersecurity agency** can not only be used to fight the Gamut infection, but it may also prove useful in defending against other cybersecurity threats. This is a positive externality since it provides benefits for stakeholders who are not part of the cooperation network. On, the other hand, because of new regulations that new companies will have to comply with, the entry barriers for new companies are expected to be significantly higher (Thorelli,1986). Furthermore, companies not in the network will be targeted more by adversaries, owing to the fact that they seem more vulnerable as they do not participate in the cooperation network. Consequently, this leads to a negative externality for the companies which do not participate in the network.

As a positive externality, the **IT company**'s spam filter will not only prevent new infections in its own network, it will also prevent currently infected machines from sending spam from the network. That is, the spam filters will indirectly protect other networks from being infected. In this sense, the spam filter can be seen as a public good, since it is non-excludable as it is impossible for the investor to exclude other parties from benefiting and use of a certain individual does not limit the availability of the product. Furthermore, spam filters are an effective countermeasure for all sorts of malware, not only the Gamut botnet. On the other side, spammers will turn their attention to worse-protected networks which will as such be infected more. This is a negative externality for parties who own these networks.

When a **citizen** increases the security on their own machine, they will no longer contribute to the botnet and the threat the botnet poses will therefore decrease. While it was not the goal of the citizen, the volume of spam will slightly decrease as a result of the improved security. This imposes a positive externality for external parties being harmed by the botnet since the total amount of spam has been decreased. However, by adopting a specific type of antivirus, you become part of the reason that malware developers will try to circumvent the protection from that specific antivirus vendor. This imposes a positive externality for external parties being harmed by the botnet since the total amount of spam has been decreased.

3 Analysis

3.1 Metric

The metric that we attempt to predict using other factors will be the Gamut infection rate per country, which was introduced in this report's Introduction. This infection rate is calculated as the number of Gamut infections within a country normalised by the number of Internet users within that respective country.

3.2 Analysis

For our analysis, we will look at what factors influence the Gamut infection rate of a country. In particular, we have gathered various factors that are plausibly related to the infection rate and look at whether there is a correlation between the factors and the infection rate. After exploring the factors, we create a linear model of the factors to approximate the infection rate.

3.3 Factors

We consider the following factors, each of which is measured at the country level, and briefly explain our motivation for including it:

1. *Computers per Capita (CpC)*[3]
   More computers mean more places to infect.
2. *Computer Science Paper Ratio (CSPR)*[4]
   An active Computer Science research community could mean better protection against botnets.
3. *Global Cybersecurity Index (GCI)*[5]
   A higher GCI could mean more mature defences against botnets.
4. *GDP per Capita (GDPpC)*[6]
   A higher GDP means there's more to lose, so there is more reason to invest in cybersecurity.
5. *ICT Development Index (IDI)*[7]
   A more mature infrastructure could mean better protection, though it could also make it easier for Gamut to spread.

[3] http://www.nationmaster.com/country-info/stats/Media/Personal-computers/Per-capita

[4] https://www.scimagojr.com/countryrank.php

[5] https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2017-PDF-E.pdf

[6] https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?view=map

[7] https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2017/MISR2017_Volume1.pdf

6. *Technology Index (TechI)*[8]
   A higher TI could mean a higher awareness of threats and thereby better protection against botnets.
7. *Terrorism Index (TerI)*[9]
   More terrorism could mean more cyberterrorism.
8. *Youth Unemployment Rate (YUR)*[10]
   More youth unemployment could mean more “script kiddies”.

Before we look at how these factors correlate with the infection rate, we need to find out how these factors correlate with each other—if two factors correlate, one of them is redundant and should be removed from further analysis. Because we will be creating a linear model, we determine the correlation as the Pearson coefficient.

	CpC	**CSPR**	**GCI**	**GDPpC**	**IDI**	**TechI**	**TerI**	**YUR**
CpC	1.00	-0.31	0.67	0.92	0.80	0.82	-0.01	-0.44
RCSP	-0.31	1.00	-0.34	-0.32	-0.61	-0.38	0.18	0.11
GCI	0.67	-0.34	1.00	0.61	0.63	0.61	0.28	-0.27
GDPpC	0.92	-0.32	0.61	1.00	0.78	0.77	-0.02	-0.39
IDI	0.80	-0.61	0.63	0.78	1.00	0.82	-0.21	-0.36
TechI	0.82	-0.38	0.61	0.77	0.82	1.00	-0.06	-0.38
TerI	-0.01	0.18	0.28	-0.02	-0.21	-0.06	1.00	0.20
YUR	-0.44	0.11	-0.27	-0.39	-0.36	-0.38	0.20	1.00

Table 1. Pairwise correlations of the factors

As shown in Table 1, there are four factors that are strongly correlated ($p \geq 0.75$ or $p \leq -0.75$) with each other: Computers per Capita, GDP per Capita, ICT Development Index, and Technology Index. We need only one of these factors, so we choose the ICT Development Index factor because it is most closely related to our problem of determining infection rate.

3.4 Linear model

When performing a linear regression, there are two hypotheses associated with each variable: a null hypothesis and an alternative hypothesis. The two hypotheses are as follows:
H0: There is no correlation between the variables.
H1: The coefficients that describe the correlation between the variables are not equal to zero.

[8] http://www.nationmaster.com/country-info/stats/Economy/Technology-index
[9] https://tradingeconomics.com/country-list/terrorism-index
[10] https://tradingeconomics.com/country-list/youth-unemployment-rate

The aim of our analysis is to determine for which factors we can reject the null hypothesis and for which it holds true.

Using R's `lm` function[11] to create a linear model, we obtain a coefficient for each factor. The linear model is shown in Table 2.

Factor	**Coefficient**	**t-value**	**Pr(>|t|)**
(Intercept)[12]	3.568e-03	3.462	0.00115
Computer Science Paper Ratio	-2.154e-06	-1.916	0.06141
Global Cybersecurity Index	-1.373e-03	-1.102	0.27595
ICT Development Index	-2.543e-04	-1.705	0.09477
Terrorism Index	-2.394e-05	-0.306	0.76075
Youth Unemployment Rate	-3.297e-05	-2.114	0.03983

Table 2. Factor coefficients in the first model.

However, linear regression can be applied to any two variables to obtain a linear model, but that does not mean that the model is statistically significant. While the p-value of the model is 0.04896 (which makes it just about statistically significant), some of the coefficients have very high t-probabilities. In particular, the t-probability of the Terrorism Index indicates that its coefficient may vary wildly. Therefore, we create a new linear model without the terrorism index.

The second model, shown in Table 3, has a p-value of 0.02597, which is better than the initial model. This time around, the Global Cybersecurity Index has a very high t-probability, so we create a new model without it.

Factor	**Coefficient**	**t-value**	**Pr(>|t|)**
(Intercept)	3.443e-03	3.498	0.000982
Computer Science Paper Ratio	-2.170e-06	-1.993	0.051638
Global Cybersecurity Index	-1.188e-03	-1.278	0.207041
ICT Development Index	-2.598e-04	-2.000	0.050850
Youth Unemployment Rate	-3.443e-05	-2.367	0.021749

Table 3. Factor coefficients in the second model.

[11] https://www.rdocumentation.org/packages/stats/versions/3.5.1/topics/lm

[12] The intercept is the constant value in the linear model and describes the noise, randomness, and errors in the model.

The final model, shown in Table 4, has a p-value of 0.02117 and each of the variables has a low t-probability, indicating that all variables are statistically significant.

Factor	**Coefficient**	**t-value**	**Pr(>|t|)**
(Intercept)	3.322e-03	3.405	0.00127
Computer Science Paper Ratio	-2.343e-06	-2.176	0.03402
ICT Development Index	-3.439e-04	-3.086	0.00322
Youth Unemployment Rate	-3.278e-05	-2.275	0.02695

Table 4. Factor coefficients in the third and final model.

By conducting the statistical analysis on the dataset, we have found which variables can be considered to have a relationship to the Gamut infection rate of a country. Conclusively, the null hypothesis can be rejected for the variables Computer Science Paper Ratio, ICT Development Index and the Youth Unemployment Rate. When creating a linear model for the infection rate based on these variables, we can say that the linear model is statistically significant because the p-value is less than the statistically significant level (which we decided to be 0.05). The model is now considered to adequately predict (to some extent of course) the infection rate based on the three known variables.

The Computer Science Paper Ratio and ICT Development Index both have a negative coefficient, which indicates that a higher value indicates fewer Gamut infections, as expected. On the other hand, the Youth Unemployment Rate negatively contributes to our linear model, even though we predicted that a higher Youth Unemployment Rate would *increase* the infection rate. A clear reason for this is hard to find, which makes it difficult to argue in favour of a causal relationship between our model and the Gamut infection rate, even if it is a statistically significant predictor.

4 Conclusion

In this report, the previously described Spamhaus dataset has been used to identify which externalities can be expected when mitigating the risk imposed by the Gamut botnet. We defined three actors who are in a position to influence the threat level imposed by the Gamut botnet and we proposed three countermeasures that could be considered. The actors we defined are the cybersecurity agency in Vietnam, a Vietnamese IT company (VPNT), and the citizens of Vietnam owning a computer.

The countermeasures we suggested entail setting up a collaboration network for ISPs, implementing spam filters and installing an anti-virus program, respectively. Consequently, there exist both positive and negative externalities related to the implementation of the countermeasures. The realisation of a collaboration network would also prove useful in combating other cybersecurity threats, but it might increase the entry barriers for new companies since they must comply with industry regulations. Installing spam filters in huge IT companies will indirectly protect other networks from being infected, however, it is also efficiently reducing the threat of all other sorts of malware. Finally, if citizens install anti-virus, they will contribute to the reduction of spam. We also claim that each of the actors has real incentives to implement the countermeasures.

Finally, we looked at factors that could possibly explain the variance in the Gamut infection rate, and then we performed a linear regression based on the factors proved to be statistically significant. Of all the factors considered, only Computer Science Paper Ratio, ICT Development Index and the Youth Unemployment Rate are considered to be statistically significant, hence the linear model is based solely on these variables. The resulting linear model serves as a prediction tool for deciding the Gamut infection rate of a country, even though it does not clearly indicate any causal relationship between the factors and the Gamut infection rate.

References

Dahlman, C. J. (1979). The problem of externality. *The journal of law and economics*, *22*(1), 141-162.

Hachem, N., Mustapha, Y. B., Granadillo, G. & Debar, H. (2011). Botnets: lifecycle and taxonomy. In *Network and Information Systems Security (SAR-SSI), 2011 Conference on* (pp. 1-8). IEEE.

Thorelli, H. B. (1986). Networks: between markets and hierarchies. *Strategic management journal*, *7*(1), 37-51.