Block 4 - Market Failures and Policy Interventions
WM0824TU Economics of Cyber Security
Group 10 – Facebook threats

Arnoud Toering - 4380061
Yorick de Boer - 4287304
Valentijn Grapperhaus - 4250338
Martine Romer - 4248759

Introduction

This assignment focusses on market failures and policy interventions. A group of actors will be analyzed for the security issue. The security issue, as described in the previous assignments, involves compromised websites. For every actor, a countermeasure will be described, including its costs and benefits, as well as possible incentives for taking action. Also, the role of externalities for the security issue will be discussed. Furthermore, a statistical analysis will be conducted. The knowledge on the security issue can be further advanced by improving the understanding of the security performance of the hosting companies. The main metric for their security performance, as described in previous assignments, is the amount of compromised websites normalized over the total amount of hosted websites. By focussing on active hosting companies per country, the impact of the economic position and ICT performance of the country on the metric is investigated. The results will provide additional insights into the underlying causes of a particular security performance, further illustrating the complexity of the security issue.

The potential role of stakeholders in the security issue: countermeasures

The hosting companies, as well as the clients of hosting companies and the individual website visitors, are very much involved in the security issue. For each of these actors, a countermeasure will be described.

With regards to the hosting companies, they would be able to mitigate the risks associated with the security issue by backing up and restoring compromised websites. In this way, the initial consequences are quite soon changed back again.
Secondly, the clients of hosting companies would be able to monitor the websites in order to detect malicious activities. Broadly speaking, this would entail that the client has increased its knowledge on the level of cyber security of its web address.
Thirdly, the individual website visitors would be able to make us of preventive measures in order to mitigate the security issue. This could be done on a general level by adopting a more alert attitude, and installing antivirus software.

Countermeasures: distribution of costs and benefits

For the countermeasures as described above, the distribution of costs and benefits will now be discussed. These distributions are perhaps best visualized using the Gordon-Loeb model (see Figure 1). as it is an important analytical model in the economics of cyber security (Gordon, 2002). The Gordon-Loeb model can be used to visualize the distribution of costs and benefits for the stakeholders, for when the countermeasure is implemented. In this way, the benefit is assumed to be the decreased probability of a successful attack.

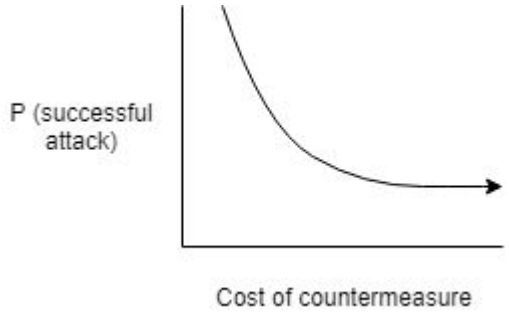

Figure 1: Gordon-Loeb model

For the hosting companies, backing up and restoring the compromised websites would lead to additional costs in a number of areas. Perhaps the most obvious one is the additional server capacity. Though it is very difficult to estimate the values for the graph, especially the value of the secured assets, the security investment must not exceed 37% of the expected loss without the countermeasure. The attack probability can be estimated on 5 percent, while the probability of occurrence was computed in the previous assignment, leading to a number of 0,002 percent.

For the clients of hosting companies, monitoring its websites would lead to additional costs in software, as well as productivity loss. Using again the Gordon-Loeb model, the optimal point can be found by estimating the value of the protected assets, as well as the attack probability and the probability of occurrence. The latter two are already given in the preceding paragraph, while the first remains difficult to determine.

Lastly, for the individual visitors of the websites, making use of preventive measures would have similar consequences as for the clients of hosting companies. Namely, it would lead to additional costs in software, as well as a loss of productivity. Depending on the transaction costs made by the visitor in order to protect itself further against compromises, the optimum point in the Gordon-Loeb model can be determined.

Incentives of involved actors to take countermeasures
For each of the actors, incentives for taking action will be discussed.

With regards to the hosting companies, this actor has a high interest in mitigating the risks and avoiding negative impact. When the compromised website owners experience damage to hard- or software, the hosting company will be responsible for the damage repayment. The amount of money will be dependent on damage, the insurance policy in place and the contract. When websites are compromised, the website owner could decide to leave the hosting company. This will result in a direct loss of customers. Furthermore, negative activities, such as hardware damage, will affect the reputation of the hosting company, leading to a decrease of new customers. Therefore, the hosting companies have an incentivize to introduce countermeasures that could mitigate the negative impact.
Secondly, the clients of hosting companies. This actor has similar incentives as the hosting company itself, namely loss of reputation and loss of customers. In this case customer refers to an individual website user. Therefore, the client also have an incentivize to introduce countermeasures that could mitigate the negative impact.

Thirdly, individual website users. This actor has an incentive to be careful, as there is no way of going back once its data is compromised. Once the data is out there, as described in previous assignments, is it very difficult to repair the damage. Therefore, this actor also has an incentive to use countermeasures to mitigate the security issue, in particular preventative measures.

Externalities surrounding the security issue

This section will briefly describe the role of externalities in the security issue. Externalities are unintended consequences, affecting those outside the initial system boundaries. As such, these consequences could have either positive or negative effects for parties that are not directly involved in the security issue. Since the security issue of these assignments is relatively broad, the externalities are derived from the same level of abstraction.

First of all, the positive externalities. When hosting companies find themselves in economically less sustainable situations due to bad reputations, this may further incentivize websites visitors to increase their online awareness. This could lead to a stimulance of their offline activities, such as shopping in physical shops. Increased online awareness could also result in better protected website visitors, which could in turn be beneficial for hosting companies as the website owners.

Secondly, the negative externalities. When cybercrime is punished relatively hard, it may result in a shift towards another type of crimes such as theft. Furthermore, when this security issue is mitigated, it will incentive attackers to look for other vulnerabilities of the hosting companies.

Statistical analysis

This section will present the statistical tests conducted for this assignment. As previously described, the tests are meant to better understand the security performance of hosting companies. A brief description of our data preparation can be found in Appendix I, as well as additional descriptive analyses of the data.

The security performance of hosting companies is visible in the metric we initially selected, namely the amount of compromised websites normalized over the total amount of hosted websites. The focus of these tests is on active hosting companies per country. This allows for an investigation on the variance in the metric among countries. Since we do not have access to the location of the individual websites, the country the hosting provider registered its AS number in is used. The new security metric therefore becomes compromised websites per country divided by total amount of hosted websites hosted by country.

Using the country level perspective, two relevant factors for statistical analysis were identified. First, the GDP per capita Purchasing Power Parity (PPP). The related dataset was obtained from *the WorldBank*[1]. This factor is chosen because it indicates the economic position of a country. When a country has a high GDP per capita there is a lot of economic activity involved. As protective measures should be in ratio to the level of economic activity, the higher the economic activity the more protective measures are expected to be in place. This in turn could influence the amount of compromised websites per country. Secondly, the *ICT development index*[2](IDI) is taken into account. This index is published by the United Nations International Telecommunication Union and is used as a tool to indicate and compare ICT performance within and across countries. This factor has been taken into account because it gives insight in the ICT performance of a country, which could influence the amount of compromised websites. Countries with high ICT performance (high IDI score) are expected to have better countermeasures in place, possibly resulting in less compromised websites.

Twice, a Spearman's rank correlation test will be conducted. This test was chosen because the relationships are not assumed to be lineair, while the the factors are not assumed to be normally distributed. The hypotheses that are posed in both tests are the following. The H0 hypothesis states that there is no relationship between the security metric and the chosen factor, so GDP per PPP or IDI. The H1 hypothesis states that there is indeed a relationship. In order to test for these hypotheses, the tests are conducted two sided. In the next section the results will be discussed.

Results

This section is meant to present the results of the statistical tests. The means of documentation is based on the guidelines as prescribed by Molin (2015).

With regards to the first statistical test, concerning the GDP per capita PPP, the results show that there is no significant relationship found between between this factor and the security metric (see Table 1). The same applies to the ICT development index. In both cases, the

[1] https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD
[2] http://apps.who.int/gho/data/node.goe.itu-idi?lang=en

correlation coefficient is relative small, while the p values are very high. Since the p values need to be smaller than 0,05 for the results to be considered significant, the H1 hypothesis is rejected for both tests.

	Compromised websites per country divided by total amount of hosted websites hosted by country
GDP per capita PPP	0,07 (p=0,721)
ICT development index (IDI)	0,12 (p=0,500)

Table 1: Correlation between security metric and GDP per capita PPP and IDI

Conclusion

In total, three actors and their possible countermeasures have been analyzed. For these countermeasures, the distribution of cost and benefits has been explained using the Gordon-Loeb model. This was followed by an explanation of their incentives, as well as a consideration of the externalities for the entire security issue. Lastly, a statistical analysis was conducted in order to determine to which extend the GDP and IDI are able to explain the variance in the security metric. From the analysis is was shown that there is no significant relationship between the security metric and the GDP per capita Purchasing Power Parity (PPP) and IDI. The found insignificant relationships could be partially explained by considering the small variance in the used data. This makes the finding of weak correlation more likely.

For future research it is recommended to focus on other factors deemed relevant for the security issue, such as updates of software, firewalls and antivirus programs. Also, future work may be increasingly benefit from bigger data sets. For these tests, we only had data for half a year of compromised websites. Bigger datasets would help to find more significant relationships and patterns over time.

References

Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)*, *5*(4), 438-457.

Molin, E. (2015). Data-analyse voor dummies. Een praktische inleiding in het analyseren en rapporteren van de samenhang tussen twee variabelen. *TU Delft.*

Appendix I - Data preparation and descriptive analysis of main dataset

This appendix is meant to show additional information on the main used dataset.

Considering the size of the dataset, we decided to write a Python script that pulled all the AS numbers from the data. AS numbers contain important information such as the domain owner and IP address. The original dataset provided the compromised website and its associated AS number. We then created a new dataframe that counted the number of compromised websites per AS number.

However, because of the large amount of AS numbers, we decided to only take into account the AS numbers that are associated with at least 10 compromised websites. Then, with the help of python package beautifulsoup and the website ipinfo.io we did a search on all the AS numbers that were selected and found all possible IP's associated with them. This gave us a total of 1.237.547.736 IP's. Then, using python package cymruwhois we found 'hosting providers' for each AS number. Finally, to make the statistical analysis possible, we grouped the hosting providers by their country.

Below, descriptive analyses are shown to provide additional insights into the dataset.

The ratio of compromised websites per small hosting provider can be seen in Figure 2. This figure shows that Russia has the largest ratio of compromised websites in the category of 'Small Hosting Providers' where respectively the United States and Greece are ranked second and third.

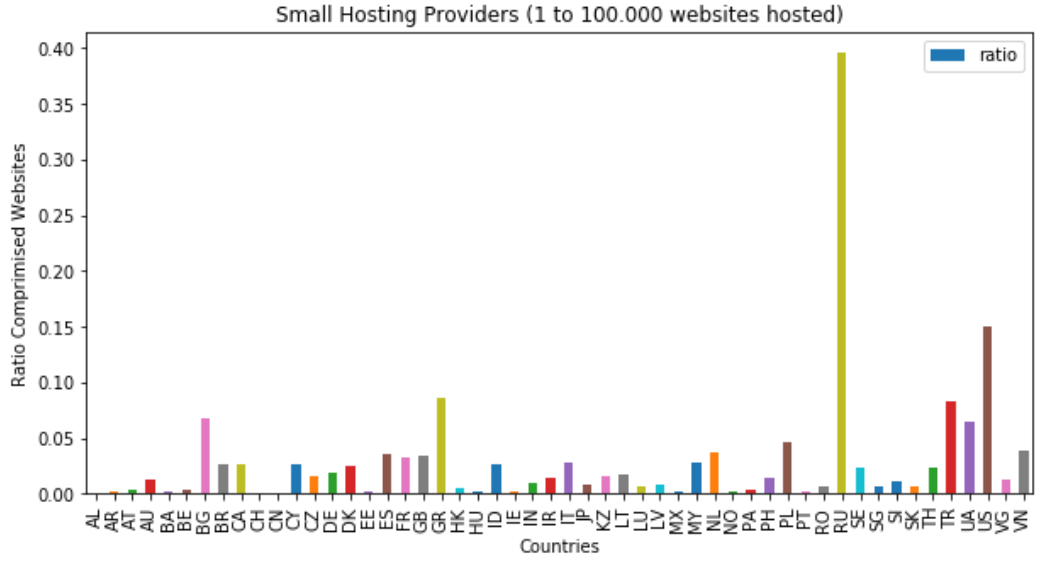

Figure 2: Ratio compromised websites for small hosting providers

The ratio of compromised websites per medium hosting provider can be seen in Figure 3. This figure shows that US has by far the largest ratio of compromised websites in the category of ‘Medium Hosting Providers’, that host 100.000-10.000.000 websites, where respectively the Germany and Great Britain are ranked second and third.

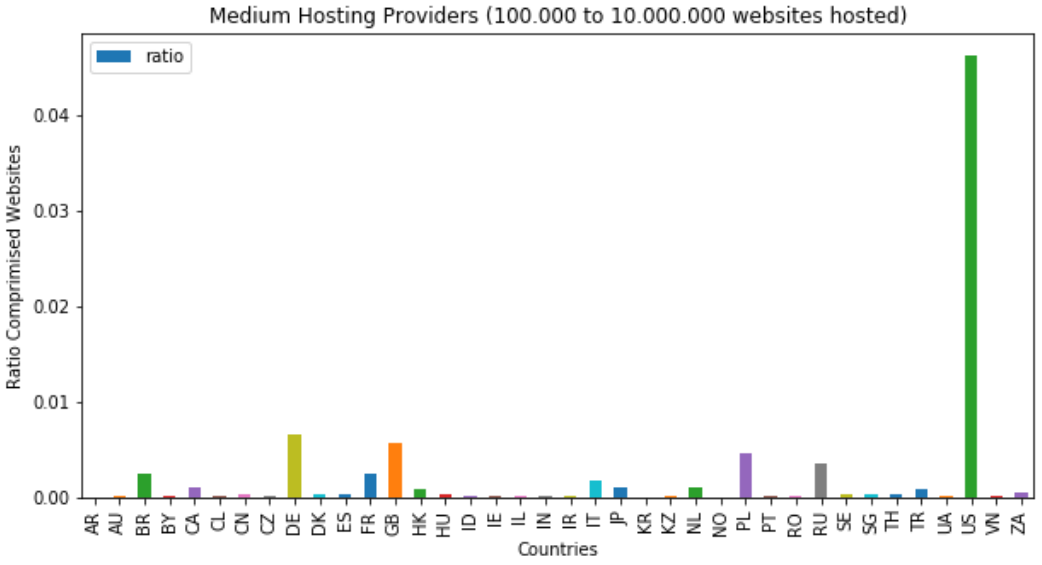

Figure 3: Ratio compromised websites for medium hosting providers

The ratio of compromised websites per large hosting provider can be seen in Figure 4. This figure shows that the US again has, by far the largest ratio of compromised websites in the category of ‘Large Hosting Providers’, that host 10.000.000 or more websites, where respectively the China and Vietnam are ranked second and third. Remarkably, in this descriptive analysis Asian countries are more present than in the previous two descriptive analyses, as shown in Figure 3 and 4.

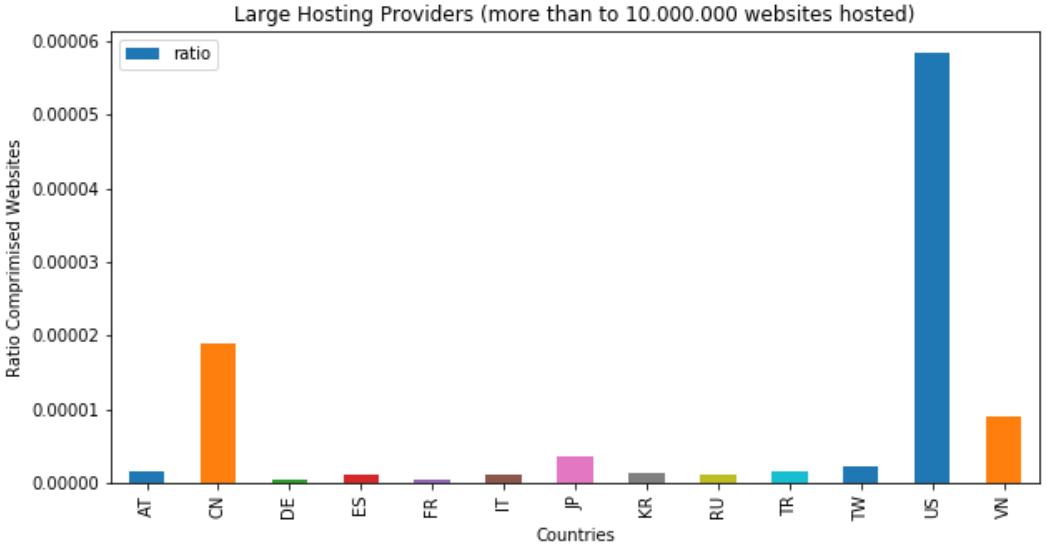

Figure 4: Ratio compromised websites for large hosting providers