Delft University of Technology

Economics of Cyber Security
WM0824TU

# Individual assignment:

# Prediction of cybersecurity level based on national cybersecurity policies

*Author:*
T.M. Nguyen (5025664)

November 15, 2019

Github repository link:
https://github.com/T-Dat/WM0824TU-Individual.git

**Abstract**

Without enough information, it is hard for governments to create effective cybersecurity policies. This paper reviews 5 high quality peer-reviewed papers to explore the difficulties governments have to face and explores the relation between cybersecurity policies and the cybersecurity level of a country. To limit the scope of the research, the cybersecurity level is limited to phishing domain activity. The cybersecurity policies are ranked using the Global cybersecurity Index, which evaluates the involvement in cybersecurity of a country. The relation is be explored with a univariate cox model to obtain the correlation between the covariate GCI and the survival rate of phishing domains. The result of the analysis shows a strong correlation that in countries with a high GCI, phishing domains are taken down faster.

## Introduction

In today's world, online services are growing exponentially. Bank clients can now manage their savings without having to leave the comfort of their houses. Online services benefit both companies and customers. Services are now more efficient, faster and cheaper. Customers have personal credentials to use these services. This, however, comes at a cost. As more and more services are digitized, the number of sensitive credentials that can be abused increases and hence it creates more opportunities for attackers to exploit these systems for personal gain. A popular attack is phishing which exploits users through social engineering. It aims to steal the valuable credentials of the victims. Depending on the critical nature of the credential, the attacker can gain access to sensible information and even perform unauthorized actions such as bank transfers.

Governments can create security policies to protect its citizen from cyberattacks. Global Cybersecurity Index (GCI) aims to "measure the commitment of countries to cybersecurity at a global level" [1]. It is a useful tool to compare countries in terms of cybersecurity. However, it does not directly measure the effectiveness of national security policies.
This study will try to provide a better understanding of the effectiveness of national security policies through the use of GCI and datasets about phishing data.

## Literature Review

Cybersecurity has to face many economic challenges with the growing rate of cybercrime. The reason for this challenge is the unbalanced distribution of responsibilities and liabilities. Organisations do not have to bear the full costs of cyberattacks directed at them. Often other parties must deal with the collateral damage. Moore gives the example of critical infrastructure where the losses are mainly incurred by society. There are three economic challenges explained in the paper: misaligned incentives, information asymmetries and externalities. [2]

Governments and companies have a different outlook on cybersecurity. Whereas the government aims to protect its citizens by creating strict security policies, companies try to maximise their benefits often at the cost of better security. These misaligned incentives make it difficult for a government to increase the national cybersecurity. For any security policies, companies will always try to find the highest return for certain security investments. Furthermore, companies can also accept the risk and rather pay a fine, because it is financially more viable. As such, there are two parties with different security intentions

at play. [2]

Organizations try to avoid sharing as much information as possible about any cyber-attacks they have been subjected to. Sharing information can cause a lot of harm to the organization. This might lead to customer and financial loss, and also make them even more susceptible to further attacks as they would appear to be easy targets. This void of data complicates the creation of proper security policies as it is hard to know what the actual security problem is and whether the situation is getting better or worse. [2]

Externalities play a dominant role in the IT industry. For example, dominant companies tend to be favoured due to the effects of the larger network. The larger the network of a product, the more valuable it is, even if there exists a superior product, but less popular. Examples of this effect are the rise of the Windows operating system, Facebook or Google. This is the reason why it is hard to introduce new products effectively. Many upgrades to security have trouble to be adopted by the community. For instance, DNSSEC and S-BGP have not been successfully adopted. This, in terms, creates a hurdle for security policies to achieve their desired goal. [2]

To solve the problems in cybersecurity, governments can establish policies to properly distribute responsibilities and liabilities to the respective parties in order to solve the aforementioned challenges. Moore reviews three solutions that apply to governments. First, ex-ante safety regulations aim to prevent accidents. These regulations oblige companies to comply with the set rules. On the other hand, ex-post liability sets fines for the responsible party in case of an accident. It tries to encourage parties to prevent these accidents by the threat of monetary fines. And last is information disclosure. Governments can impose companies to disclose about accidents and attacks. This will help to provide a better picture of the security level and hence lead to better security policies. [2]

National security policies can be considered as security investments as it requires an investment to obtain a certain benefit. Thus, security policies also benefit from investment models to determine their return. A popular choice for comparing investments is the ROSI model which computes the return on security investment. To establish a security model, the costs and the benefits need to be determined to compute the ratio. Direct costs are easily obtainable, whereas the benefits and indirect costs are not. Due to the lack of security-related data, it is more difficult to create an accurate model of the situation. Furthermore, security metrics need to be carefully chosen to provide any useful and valuable information. These problems can influence the result in such a degree that bad or less optimal investment is pursued. This issue does not facilitate the creation of good and effective security policies. [3]

One common form of cybercrime is phishing, which aims to deceive users into divulging sensitive information for financial or informational gain. Phishing is performed using methods of messaging, such as email, telephone, SMS, etc. For this study, only phishing emails will be considered. The simplicity of sending emails explains the popularity of phishing. It is easy to send tremendous amounts of phishing emails. There exist counter-measures provided by email services such as spam filters. However, emails are only one part of the attack. Links to deceptive websites are key to a successful attack. These websites are crafted by the attackers in such a way to deceive the victim to input

sensitive information. These websites need to be registered and hosted by servers for the attack to work. This is another area where countermeasures can mitigate the problem by taking malicious domains down. [4]

In the dataset used about phishing domains, their start and end time of their up-time can be found. When considering their end time as a failure, a survival curve can be derived. This allows establishing a Cox model to interpret the results. In the paper, Cox defines all the formulas and steps to obtain such a model. [5]

The Global Security Index aims to rank countries in the field of cybersecurity. This is achieved through the combination of 25 indicators to evaluate the level of commitment in cybersecurity of a country. The goal is to provide an incentive for countries to improve their ranking and thereby their national cybersecurity. The GCI exposes areas of improvement for countries. The framework is based on 5 pillars:

1. Legal
2. Technical
3. Organizational
4. Capacity building
5. Cooperation

Note that the GCI does not directly measure the cybersecurity level of a country, but the level of commitment. [1]

## Research Question, Objective and Hypothesis

There is a handful of indexes aiming to rank countries on criteria in the cybersecurity domain. Governments can evaluate their involvement in cybersecurity using these indexes. One such index is the Global Cybersecurity Index. As previously mentioned, the GCI only measures the involvement of a country in cybersecurity but not the actual state [1]. The ability to get a concrete value for the security level is a very valuable tool for governments [3]. This index can be used to check how effective implemented securities policies perform ie. how profitable is the security investment. However, such an index is very difficult to accurately compute as a lot of data would be needed. The issue is that organisations tend to share as little information as possible in regards to cybersecurity in the fear of reputation damage and other losses [2]. This leads to the research question:

Do the cybersecurity policies of a country correlate with the overall national cybersecurity level?

The research question holds two main elements: cybersecurity policies of a country and national cybersecurity level. The first can be measured using the GCI from the dataset[**?** ]. For the latter element, as there is no available metric or index, only a very specific and small subset of cybersecurity will be considered with the assumption that the subset of cybersecurity approximately reflects the overall cybersecurity. As this research paper is built upon previous assignments from the course "Economics of Cyber Security", the

already provided dataset and previously computed metric will be used. The dataset in question is the phishing dataset from Clean MX [6] and the derived metric is the survival rate of phishing domains 2. This leads to the following hypothesis:

The survival rate of phishing domains is lower in countries with a high Global Cybersecurity Index.

In the following sections, a statistical analysis will be performed to support or reject the hypothesis. The hypothesis will be accepted if the $p$-value is smaller than 0.05, otherwise, it will be rejected.

## Methodology

The hypothesis will be tested using a quantitative research approach to the problem. Two datasets [6**?** ] are used in this analysis. The programming language R will be used to perform the analysis.

To obtain the survival rate of phishing domains, the dataset [6] is used. First, failure needs to be defined. Here failure is the moment when a phishing domain is taken down. The time until failure can then be obtained $lasttime - firsttime$ (assumption: firsttime denotes the date the domain was hosted and lastime the date that the domain was taken down). The dataset is right-censored meaning that for certain entries/domains, no lasttime is given. These entries will be treated as if they were still up on the 01-01-2017. This date was selected because looking from the data, it seems that this was the date when the last entries were entered. Any entries taken prior to 01-01-2015 are removed to ensure the temporal relevance of the result. With these values, the survival rate of the phishing domains can be obtained, see 2.

| Member State | Score | Global Rank |
|---|---|---|
| United Kingdom | 0.931 | 1 |
| United States of America* | 0.926 | 2 |
| France | 0.918 | 3 |
| Lithuania | 0.908 | 4 |
| Estonia | 0.905 | 5 |
| Singapore | 0.898 | 6 |
| Spain | 0.896 | 7 |
| Malaysia | 0.893 | 8 |
| Canada* | 0.892 | 9 |
| Norway | 0.892 | 9 |
| Australia | 0.890 | 10 |
| Luxembourg | 0.886 | 11 |
| Netherlands | 0.885 | 12 |

Figure 1: Snippet of GCI dataset

The GCI dataset [1] lists the index value for each country. An extract can be seen above in figure 1 with the countries with the highest GCI. As the country where the domain is hosted is given in the phishing dataset [6], each entry can be associated with the correct GCI value. A univariate Cox regression analysis is done on the survival rate of the phishing domains and the covariate GCI to obtain a Cox model. Then, the proportional hazard assumption is tested to see whether the obtained Cox model correctly models the relationship between the covariate and the dependent variable. [5]

## Results

The result of the univariate Cox regression analysis between the GCI and the survival rate can be found below in 1. The $p$-vale found under $\Pr(>|z|)$ of <2e-16 is very small indicating the global statistical significance of the obtained model. Next, the result of the Wald test under z of 9.368 shows that the variable GCI significantly helps to predict the dependent variable. The two values indicate that the obtained results for the regression coefficient, the hazard ratio and the confidence interval are significant to the Cox model. The positive regression coefficient under coef of 1.9097 shows that the hazard is higher for countries with GCI. At first glance, this might seem contradictory. However, hazard, in this case, is a desirable outcome as it is defined as "phishing domain taken down". The considerable hazard ratio indicates that for an increase of one unit in the GCI, the hazard is increased by 657 % within an upper and lower 95 % confidence interval of

[452.7%, 100.7%].

| | coef | exp(coef) | se(coef) | z | Pr(>\|z\|) | exp(-coef) | lower .95 | upper .95 |
|---|---|---|---|---|---|---|---|---|
| **GCI** | 1.9097 | 6.7508 | 0.2039 | 9.368 | <2e-16 | 0.1481 | 4.527 | 10.07 |

Table 1: Cox model results

The Cox model only holds if the proportional hazard assumption holds. The result of this test are given below 2. Since the $p$-value of 0.222 is significantly greater than 0.05, there are no time-dependent variables. Thus, the proportional hazard assumption holds. Hence, the obtained Cox model properly models the regression.

| | rho | chisq | p |
|---|---|---|---|
| **GCI** | 0.0118 | 1.49 | 0.222 |

Table 2: Proportionality hazard assumption test

The obtained results show that there is a significant correlation between the survival rate of phishing domains and the GCI of the countries where they are hosted. The positive regression coefficient indicates that the hypothesis hold. The survival rate of phishing domains is lower in countries with a high Global Cybersecurity Index. With the two established assumptions, the research question can be answered. Cybersecurity policies of a country correlate with the overall national cybersecurity level. However, these assumptions might lead to wrong conclusions. They will be further discussed in the following section.

Even though measuring the involvement in cybersecurity (GCI) is not equal to measuring cybersecurity levels, the GCI is still a good indicator for evaluating the cybersecurity levels. The positive correlation found in the analysis indicates that investment in cybersecurity policies improves overall security. Policies create the right incentives for organisations to improve their security and thus contribute to the national level.

## Limitations

There are several limitations that might have affected the effectiveness of the study. A high percentage of the dataset comes from a select number of countries and from bigger companies. Countries with a lot of entries have too much weight on the end result. Countries with a small number of entries might lead to wrong conclusions as it is not unlikely that the result is due to anomalies. Furthermore, only phishing domains that have been caught are present in the dataset. It might be that there is a significant number of phishing domains that are effective at evading detection and thus are not present in the dataset. The dataset is only a small subset of the overall population of phishing domain. It is not possible to know whether this subset represents a valid random sampling. The research could be improved by confirming the results with phishing datasets coming from different sources.

Another limitation is the GCI. The GCI is determined by a survey composed of 150 questions. The accuracy of the GCI depends on the willingness of governments to share all the information and on the accuracy. Furthermore, the survey might not encompass all the factors in the cybersecurity policies. As with the phishing dataset, the result of the

research could be confirmed with another index that also aims to measure cybersecurity policies.

Next is the assumption that the state of phishing somewhat reflects the overall cybersecurity state. To improve the research, the correlation with the GCI should be done with additional data of different security areas.

For the dataset used about phishing domains, some assumptions had to be made for the right-censored data, as no explanation was given on the entries. There is a possibility that the data was misinterpreted resulting in an incorrect survival curve.

The results of the analysis only show correlation, not causation. The result does not give factors that might explain the correlation. The found correlation from the analysis might be spurious due to chance. The causation between the GCI and the cybersecurity level could be investigated in the future to confirm the relationship.

## Conclusions

With the growing e-commerce, there is an increasing need for better cybersecurity to protect the customers as well as organizations. Governments create policies to push the digitization in the direction of a safer cyber environment where organisations and users can interact without any fear. To create appropriate policies, governments need the proper metrics to evaluate the situation. Only a few of such metrics are available such as the Global Cybersecurity Index (GCI) [1]. However, none of them measure the cybersecurity level of a country but only its involvement in cybersecurity ie. cybersecurity policies. The research question that was derived from the security issue is as follows: Do the cybersecurity policies of a country correlate with the overall national cybersecurity level? Measuring cybersecurity level can pose a challenge for this research. Instead of focusing on cybersecurity as a whole, the hypothesis reduces the scope only to phishing [4]. A Cox model [5] has been created using the GCI and phishing data. The analysis concludes that there is a high correlation between the two variables and thus supports the hypothesis derived from the research question. This, in turn, enables to answer the research question. The strong correlation shows that security policies do in fact correlate with the cybersecurity level.

Governments can create better security policies due to the ability to evaluate the national cybersecurity level effectively using the GCI. The insight gained helps to create better policies. It allows to combat to a certain degree the economic challenges faced by cybersecurity in Moore's paper [5]. The new policies can help to decrease information asymmetry, limit externalities and create the right incentives to lead the national security level in the right direction. Furthermore, the GCI can be used in hindsight to evaluate the policies and potentially improve them or create better ones.

# Appendix

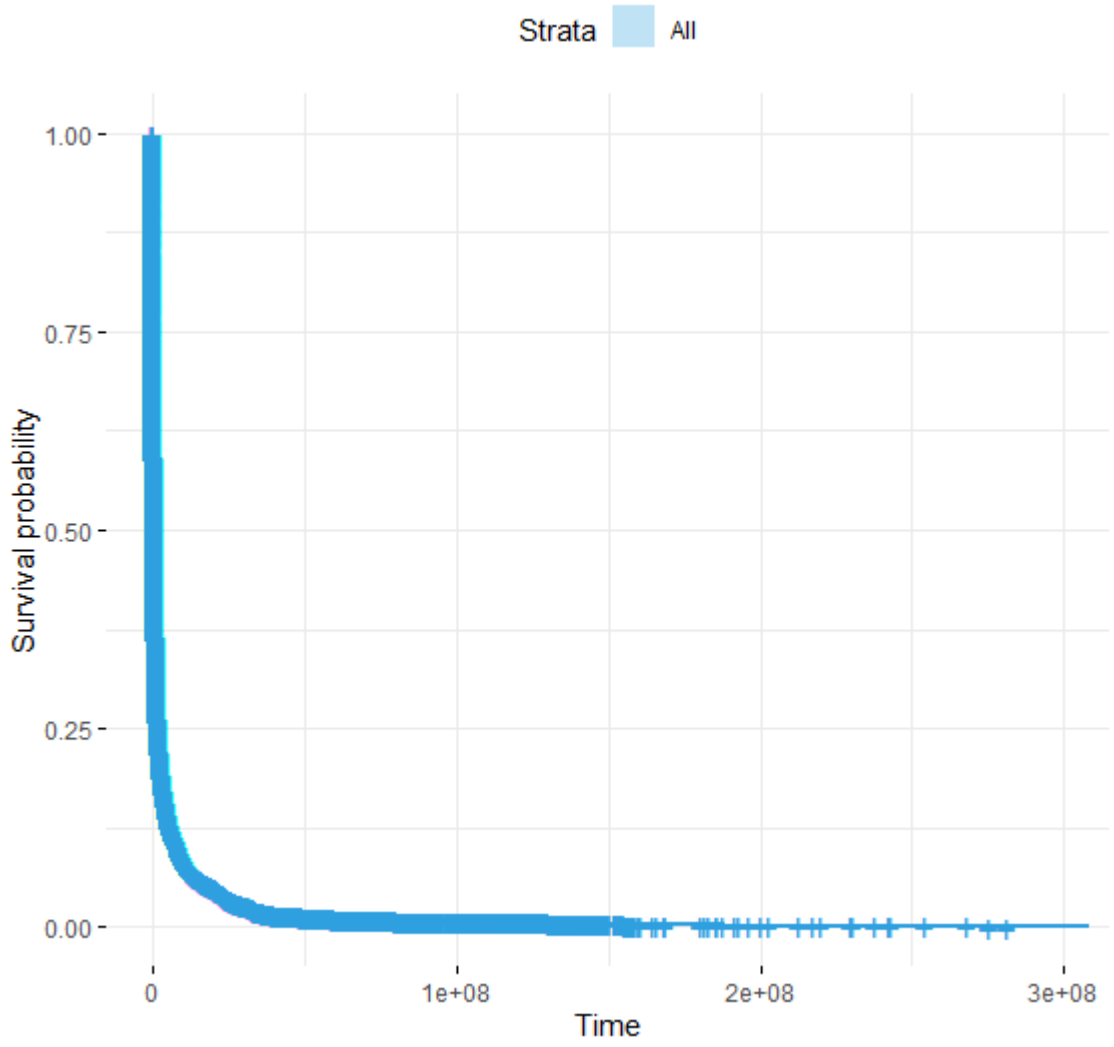


Figure 2: Survival rate of phishing domains

# References


ITU, "Global cybersecurity index (cgi) 2018," 2018.

T. Moore, "The economics of cybersecurity: Principles and policy options," 2010.

R. Böhme, "Security metrics and security investment models," 2010.

J. Chaudhry, S. Chaudhry, and R. Rittenhouse, "Phishing attacks and defenses," vol. 10, pp. 247–256, 01 2016.

D. R. Cox, "Regression models and life-tables," 1972.

C. Mx. phishing. [Online]. Available: https://support.clean-mx.com/clean-mx/phishing.php