Economics of Security — Market Failures and Policy Interventions

Tycho Teesselink (s1560506)
Tariq Bontekoe (s1451278)
Æde Symen Hoekstra (s1479679)
Ramon Houtsma (s1245228)
Group 13 — Phishing

October 2018

1 Introduction

The security issue that this document covers is phishing. We have a dataset that consists of phishing website records for different companies, e.g. PayPal. We have created a metric that determines how well a company performs on the security level by comparing uptime of a phishing website. An example of this can be seen in Figure 1 for three companies in the dataset. We created this graph by doing a survival analysis [1] per company. Here, the lower the (survival) curve, the better [2]. This metric will be used again later on, to determine what factors play a role in the security performance of these companies. The survival curve tells which ratio of the phishing sites is still up after a certain amount of days.

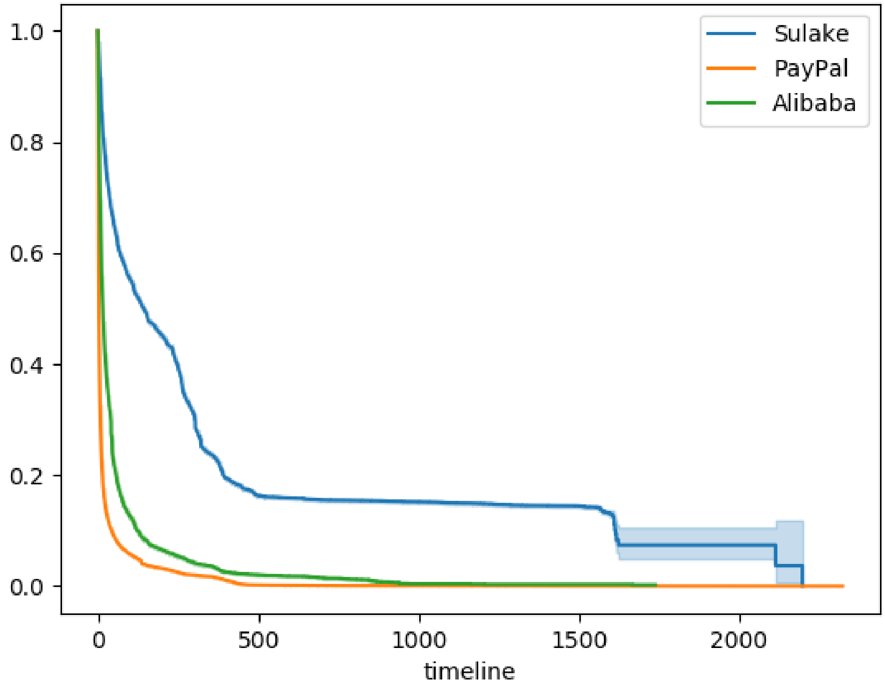


Figure 1: Survival curves for Sulake (creator of the popular game ‘Habbo’ [3]), PayPal and Alibaba with uptime in days on the x-axis and ratio of phishing sites still up on the y-axis. The shaded area is the 95% confidence interval.

2 Actors involved in the Security Issue

In this section we describe what measures three main actors, namely 'Online Service Providers', 'Hosting Providers' and 'Online Service Consumers', could take to mitigate the security issue. For each countermeasure we discuss what effect implementing this countermeasure might have on other actors involved.

2.1 Online Service Providers

When an online service provider is an actor to be targeted by phishing, we mean that their service is attractive to abuse. For example PayPal is an online service provider. When PayPal login information is stolen through phishing their service will be used to transfer money which leads to losses.

2.1.1 Countermeasure

Financial services can buy and install software that tracks payments. Anomalous payments due to stolen login information can be detected with this software and therefore these 'malicious' transactions can be halted before they are processed.

2.1.2 Distribution of Costs and Benefits among the Actors

- **Online Service Providers:** Service providers will benefit by having fewer fraud cases associated with their services. Furthermore, they save money because of fewer reimbursements to clients. There are costs involved for buying and maintaining fraud detection software, however the trade-off with less overhead with fraud cases and improved reputation for reliability is worth it.
- **Hosting Providers:** Hosting providers are not affected by this countermeasure, since the countermeasure does not target hosted phishing sites.
- **Online Service Consumers:** Consumers will probably not notice anything, even though they have a lot of benefit by this solution. Some of the costs of this solution might be written off to its service consumers, however this will be relatively little compared to the standard costs. They do have a lot of benefit by this solution however, as for example transactions done when their credentials are stolen may be halted.

2.1.3 Incentive Analysis

Online Service Providers of financial services definitely have incentive to implement this countermeasure. The number of fraud cases will go down, which leads to less overhead with fraud cases, fewer reimbursements, improved reputation for reliability and higher customer satisfaction. Online Service Providers of other types of services, e.g. e-mail services or social networks, will have less incentive to implement fraud detection in transactions. It depends on whether they handle a lot of transactions by customers or not. E-mail providers will benefit the least because they are free or work with subscriptions. Social networks may have some incentive if they use a virtual currency within their network which users can buy for in-game purchases for example.

2.1.4 Externality Reflection

Implementation of this countermeasure by an online service provider may affect customers and competitors:

- **Customers:** Customers of the online service provider receive positive externalities in the form of a smaller chance of being a victim of fraud with the service. A negative side-affect may be that their data is being gathered. If a data breach were to happen, this personal data might become known.
- **Competitors:** Competitors of the online service provider may lose customers to the service provider that implemented the countermeasure, because of the improved perceived service and reputation that are the result of the countermeasure.

2.2 Hosting Providers

The hosting providers as an actor are responsible for hosting the phishing sites that target customers of online services. It may be their moral responsibility to not facilitate adversaries with their services. It is a possible that phishing sites are allowed inadvertently because of insufficient background checking before assigning a domain.

2.2.1 Countermeasure

A more thorough information check before accepting webhosting applications. This means that people, including phishers, have to provide more information than for example just a name, email address and bank account. This will make it less attractive for phishers to register a hosting solution as they have to provide more information which might reveal who they are.

2.2.2 Distribution of Costs and Benefits among the Actors

- **Online Service Providers:** Online service providers will not have increased costs. They do however benefit from this countermeasure. As this measure is supposed to lower the amount of phishing websites, they will be less likely to be abused by phishing. This results in less reputational damage that may be incurred during a phishing campaign.
- **Hosting Providers:** Direct costs for hosting providers will not change much. It may be necessary for some initial development to support this change. Indirect costs may be much higher. People are usually not very keen on sharing a lot of information about themselves. Therefore if more information is required, some people may decide not to apply for a hosting solution at all. This will result in lower revenue for the hosting company itself. The benefit for hosting providers is that their reputation will

be increased. Hosting less phishing websites will show that they are an innovative and healthy organisation within the society.

- **Online Service Consumers:** Online service consumers can benefit the most from this countermeasure. The purpose of the measure is to protect the consumers by decreasing phishing. They will not see any increased costs, even if they want to get their own hosting solution. Here we assume that these people will not mind providing the extra information under the reason that 'legitimate people have nothing to hide'.

2.2.3 Incentive Analysis

Hosting providers may not have a very large incentive to implement this countermeasure. The costs and annoyances that this measure introduces will hinder some of the revenue that these companies can get. One of the main things that distinguishes one hosting provider from others is the ease of use of the platform and the customer friendliness. If such a measure as proposed above is introduced, this might compromise the quality of service.

2.2.4 Externality Reflection

This countermeasure may also impose negative externalities on the online service consumers. More personal data gathered also means more risk of a privacy breach. A positive externality may be that it is easier to catch the culprits if they still decide to get a hosting solution and host a phishing website. So for example the police may be able to apprehend them quicker.

2.3 Online Service Consumers

Consumers of a certain Online Service Providers may be targeted by Phishing Attacks, which trick them into for example revealing their login information. These credentials might give criminals access to for example a bank account. The criminals can then write off money from the consumer's account to themselves in order to make a profit. This will lead to the consumer losing money, which is rather inconvenient for the consumer.

2.3.1 Countermeasure

Consumers could participate in a phishing awareness training. This will help consumers identify phishing websites, which means they are less likely to fill in their personal information and therefore a part of this risk will be mitigated.

2.3.2 Distribution of Costs and Benefits among the Actors

- **Online Service Providers:** Unless the service provided is a phishing awareness training, this actor will not directly benefit, nor have costs for this countermeasure.
- **Hosting Providers:** Hosting providers have no costs, and no direct benefits from this countermeasure.

- **Online Service Consumers:** The online service consumers will bear the costs for this, however they also benefit the most. They will have to pay for the training as well as spend the time taking it.

2.3.3 Incentive Analysis

The online service consumers may have quite some incentive to take a phishing course. By taking such a course, and maybe spend a few dollars you gain a lot more security when browsing the web. It also may prevent users from filling in their information on a phishing site, which will help them even more. The costs that may be required should be seen as an investment rather than sunk costs. However, this may not be the case for all online service consumers. Some might only use the service once in a while, and therefore it would not be feasible for them to take such a course. It would still be useful, but the costs may outweigh the benefit for them.

2.3.4 Externality Reflection

This countermeasure has a positive effect for online service providers, because less customers will fall victim to phishing sites that target customers of their service. It might also be possible that these consumers can report fishy websites that may be used for phishing. This will help hosting providers in blocking websites more quickly.

3 Actor Security Performance

3.1 Factors causing Variance

We suspect that the variance between companies can be explained by their respective maturities in phishing site detection as well as the technological developments in the countries. The former can most likely be explained by the sector in which a company lies as companies in the financial sector have been dealing with these kind of attacks far longer than others. For the latter factor we have chosen to couple countries to continents, one of the reasons for this is that technological progress for countries belonging to the same continent is somewhat similar. Another reason is that it was too hard and elaborate to make a distinction for every country.

- **Sector:** We suspect that part of the difference in our metric is explained by the sector in which our companies work. To be more precise we expect a difference between companies from the financial sector and companies from other sectors.
- **Continent of headquarters ('residence'):** A part of the difference between companies might be explained by their country of residence, as we do not have enough data to distinguish between countries, we limited the scope to continents as we also expect to see differences there.

3.2 Data

To determine to which continent the companies belong or in which sector they belong, we performed a quick online search. We mainly used the websites of the company itself, or Wikipedia [4]. We labeled the companies accordingly to what was found and subsequently grouped the data respectively per continent or per sector. We define the continent of a company to be the continent where the main headquarters of the company resides and not as the main market of the company.

The dataset was also cleaned to the extent that most companies that are the same but have different labels are now merged in the same company. Furthermore we did not consider data points on which we could not find appropriate information in order to put them into the right category for our statistical tests. This was done in order to prevent incorrect results and could be done as there were enough data points in order to not consider a part of it.

3.3 Statistical Analysis

We

3.3.1 Sector

Because our data is mainly consisting of financial and gaming companies we have decided to split based on sector. We split on 'financial' (F), and label the rest as 'other' (O). Subsequently we performed a logrank test to determine whether financial institutions perform better in regard to taking down phishing websites [5, 6]. The way this works is we create an initial hypothesis (H_0) which says that there is no difference in performance between companies in the financial sectors and companies in other sectors. We subsequently either reject or accept this hypothesis with a confidence level of 95% if the p-value is not greater than 0.05. The survival curves belonging to each sector can be seen in Figure 2a and Figure 2b for the logarithmic scale. The survival curves in the graphs differ quiet a bit. The curve for the financial sector is much lower than the 'other' curve. This would insinuate that financial companies are performing better in regard to taking down phishing websites. This is also apparent in the p-value from the logrank test. Comparing financial against the other sectors resulted in a value of 0.0000 (4 digits precision) therefore we reject the null hypothesis which stated that the two did not differ and can hence conclude that companies in the financial sector perform better than the ones in the other sectors.

In addition to the logrank test, we used a Cox proportional hazard model to determine how much the financial sector performs better than the other sectors. This model can be used to calculate a hazard rate [7, 8]. This rate gives insight into with what percentage the hazard is reduced or increased for a certain category.

coef	exp(coef)	se(coef)	z	p	lower 0.95	upper 0.95
0.5660	1.7613	0.0059	96.5332	0.0000	0.5545	0.5775

Table 1: Cox proportional hazards model results for the financial sector (4 digits precision)

The results of this test can be found in Table 1. From these exp(coef) in results we see that the hazard in the financial sector is reduced by about 76% within the 95% confidence interval when compared to the other sectors. We can therefore conclude that the financial sector performs better than the other sectors combined. However, this does not necessarily mean that it is the best performing sector.

3.3.2 Continent

We also test whether the continent where the company is based, has influence on how quickly phishing websites are taken down. We split the data on six different continents: North-America (NA), South-America (SA), Asia (AS), Europe (EU), Africa (AF) and Oceania (OC). Initially we also had a distinction for the Middle East but later we decided to merge this with Asia. Subsequently we have performed the logrank test again on our initial hypothesis (H_0) which compares two continents. We want to compare all possibilities, therefore we perform a pairwise logrank test for all continents.

	SA	AF	EU	OC	AS
NA	0.0000	0.0000	0.0000	0.0003	0.0000
SA	🟥	0.0000	0.0000	0.0000	0.0000
AF	🟥	🟥	0.0000	0.0000	0.0000
EU	🟥	🟥	🟥	0.0000	0.0000
OC	🟥	🟥	🟥	🟥	0.0000

Table 2: p-values for the logrank test with the continents (4 digits precision)

continent	coef	exp(coef)	se(coef)	z	p	lower 0.95	upper 0.95
NA	0.4633	1.5893	0.0079	58.2887	0.0000	0.4477	0.4789
SA	0.1211	1.1288	0.0203	5.9802	0.0000	0.0814	0.1608
EU	-0.6049	0.5461	0.0095	-63.5677	0.0000	-0.6236	-0.5863
AF	0.3303	1.3914	0.0526	6.2831	0.0000	0.2273	0.4334
AS	-0.2151	0.8065	0.0195	-11.0390	0.0000	-0.2533	-0.1769
OC	0.0309	1.0314	0.0324	0.9540	0.3401	-0.0326	0.0943

Table 3: Cox proportional hazards model results for different continents (4 digits precision)

Our initial hypothesis is that there is no difference in performance between companies based in different continents. Based on the results we reject the hypothesis with a confidence level of 95% for all continents. The *p*-values (Table 2) were not high enough (> 0.0500), which means that the continent of origin for a company is a determining factor on how well they take down phishing websites. We also provided the survival curves for these continents in which you can see that there is indeed a difference between different continents, these can be seen in Figure 3a and with a logarithmic scale in Figure 3b. This test however, merely says that there is a difference between the continents, not how much of a difference and therefore which one performs better. Therefore we calculated the hazard rate for each continent when compared to the others. The results hereof can be found in Table 3, we see here that only one of the hazard rates is not significant, namely the one belonging to Oceania (OC). Thus for Oceania we do not have a statistically significant result telling us how much of the hazard is increased or decreased for companies in this category. For all the other countries the results are significant and we can see that companies in North-America, South-America, and Africa have reduced risk and companies in Europe and Asia have increased risk.

3.4 Limitations

There are some limitations to the statistical analysis performed above. One drawback is that the dataset is quite biased, a high percentage of the data comes from the same small set of companies. This leads to the drawback that some companies have a very large influence on the category the fall in. This is a rather hard problem to solve as undersampling some of these companies or adding weights might give too much influence to phishing attacks based on the smaller companies. These latter ones are more likely to be anomalies and hence give unrealistic results, because of this we decided to just take the data as is and describe this limitation.

One other possible issue is that for the Cox proportional hazard model we had to use random sampling, for otherwise the gradient descent that is used in this method was too slow. Random sampling is in principle not an issue, but might influence the results a little bit, thus we have weighed the amount of data versus the speed of convergence in order to get a result that is at least very close to the real value.

4 Conclusion

We have performed several tests on aggregated data to determine factors that influence the security performance in regard to phishing website takedown. Different continents perform differently in performance of phishing website takedown. We found that when pairwise compared, we can say with 95% confidence that all seven continents differ in performance. These results were found statistically significant. We have also found that North-America and Africa perform the best, according to the Cox proportional hazards test. The EU and Asia perform the worst. These results are comparable to the results we found in the survival curve of the continents in general. We have also performed the same tests when the data was aggregated on business sector. We made a distinction between the financial sector and others. We found that the financial sector performs much better than others. Reasons for this may be that the financial sector has been dealing with these kinds of problems for much longer than others. Another reason is that their incentive to take down phishing websites is higher, since the expected loss is higher for these types of companies.

References

[1] "Survival Analysis Basics - Easy Guides - Wiki - STHDA." [Online]. Available: http://www.sthda.com/english/wiki/survival-analysis-basics

[2] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J Am Stat Assoc*, vol. 53, pp. 457–481, 1958.

[3] "Sulake." [Online]. Available: https://www.sulake.com/about/

[4] "Wikipedia." [Online]. Available: https://www.wikipedia.org/

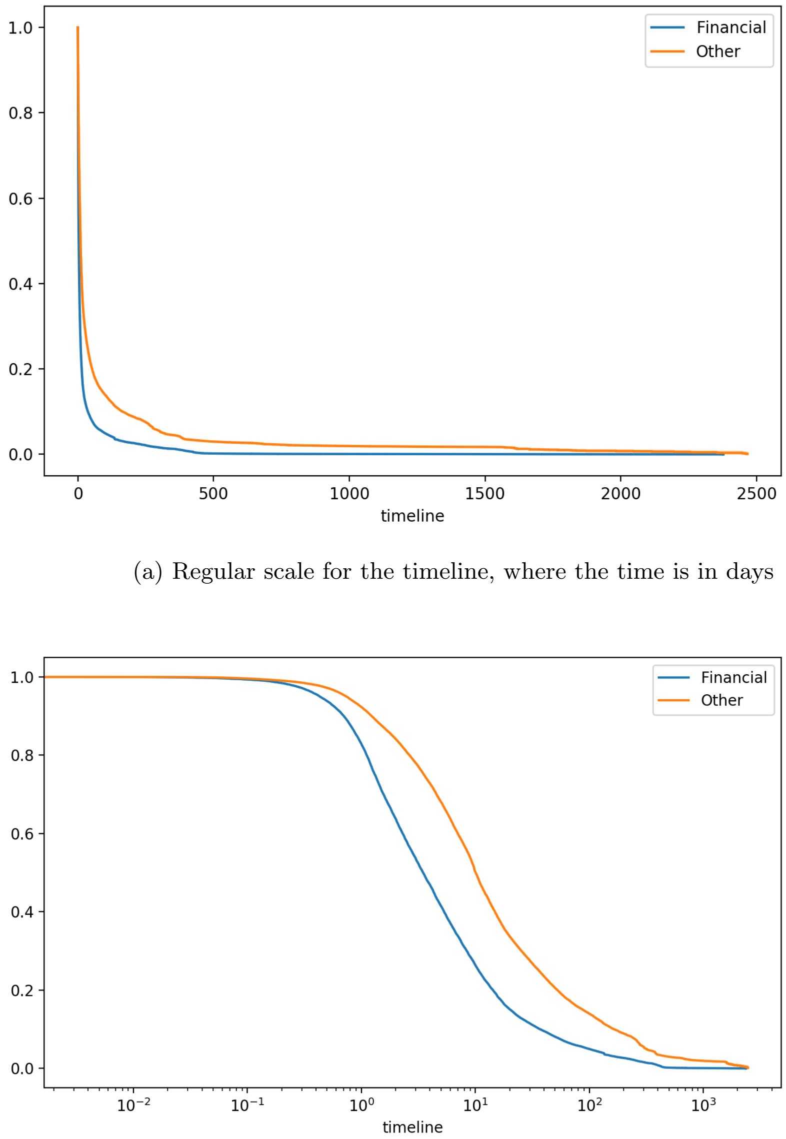
[5] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman, "Survival analysis part II: Multivarate data analysis – an introduction to cencept and methods," *British Journal of Cancer*, vol. 89, pp. 431–436, 2003.

[6] S. Pocock, T. C. Clayton, and D. G. Altman, "Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls," *Lancet*, vol. 359, pp. 1686–1689, 2002.

[7] D. R. Cox, "Regression models and life tables (with discussion)," *J R Statist Soc B*, vol. 34, pp. 187–220, 1972.

[8] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman, "Survival analysis part I: Basic concepts and first analyses," *British Journal of Cancer*, vol. 89, pp. 232–238, 2003.

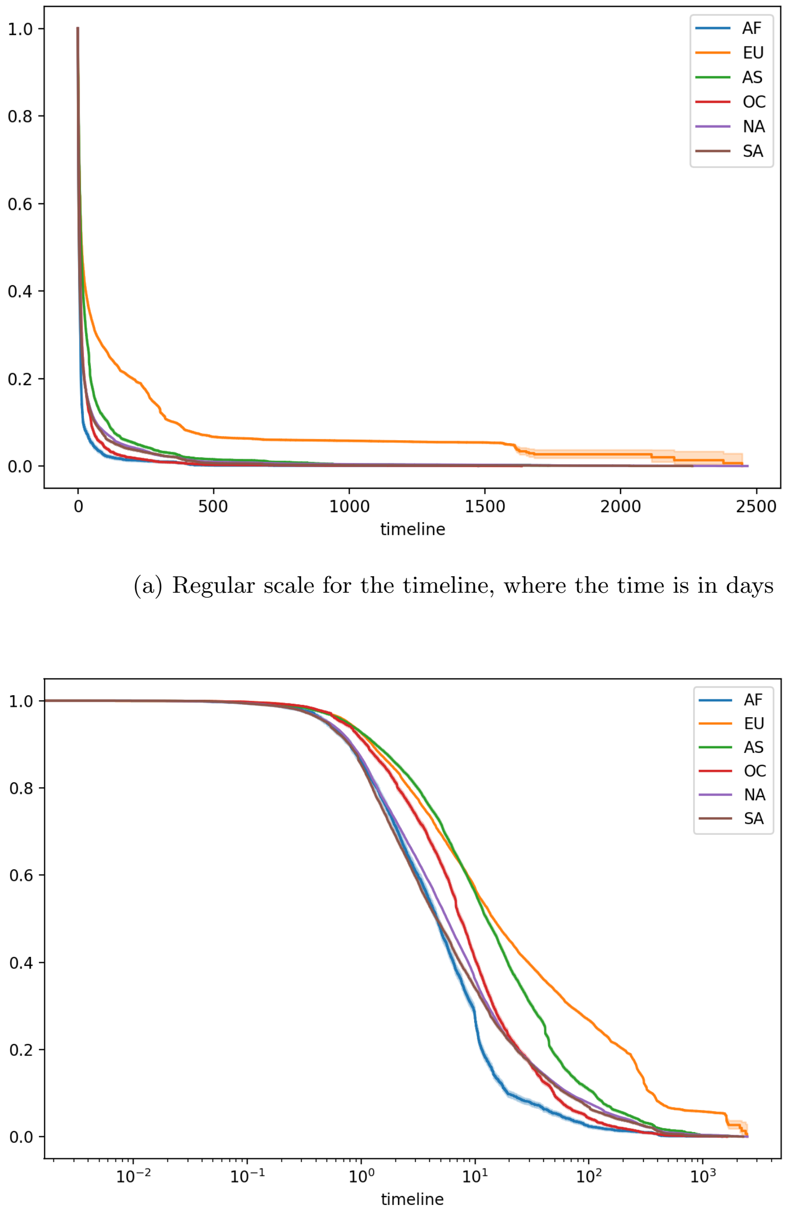
A Figures



(a) Regular scale for the timeline, where the time is in days

(b) Logarithmic scale, also time in days

Figure 2: Survival curve for companies in the financial sector compared with other sectors



(a) Regular scale for the timeline, where the time is in days

(b) Logarithmic scale, also time in days

Figure 3: Survival curve for companies originating from six different continents