Economics of Cyber Security

Peer Review Group 10 (Facebook ThreatExchange)

Marco Dijkslag	Mathay Kahraman	Yoep Kortekaas	Sam Witt
s1743716	s1724665	s1719734	s1672509
4985028	5124182	4984595	5100356

September 25, 2019

1 Summary of the assignment

The motivation for this report is the problem of phishing, which can appear in different forms. Phishing is a serious problem for the direct victim, but also for other involved parties, e.g. websites where phishing links get posted may lose their visitors. The main focus of this report is on companies/organisations whose website is crucial for business continuity. A proposed ideal security metric is determining the likelihood a posted URL is harmful based on the domain, origin of the URL, and whether the poster of the URL has posted harmful links before. An ideal metric to measure the impact of phishing is to measure the damage phishing links cause, for example the costs of the damage. This metric in combination with the metric that measures the likelihood an URL is related to phishing. If the expected damages of a phishing incident increases, the minimum likelihood required is lowered. So more links get removed from the platform. Other metrics to measure damages are the likelihood a phishing link is clicked, the amount of users targeted, and what the damage of one incident is. Lastly, an ideal metric is one that measures user behaviour and awareness around phishing attacks to determine if investment in security measures should be made. Existing metrics in practice are the metrics used by threat protector from Akamai, phishing prone percentage gathered from Phishing Security Tests (PSTs) to raise phishing awareness, the number of clicks on phishing websites, the percentage of visitors that provide personal information, and URL characteristics like the domain name that may be known for phishing attempts. Defined metrics from the provided dataset are: determining whether a URL links to a phishing website using the top-level domain, keywords in phishing links, and the number of special characters used in phishing URLs. The results of these metrics are that the most frequent phishing links use ".nf/" as top-level domain. The most occurring word inside a phishing link is 'in'. A limitation of the metric is that there is no data on non-phishing URLs, thus no comparison can be made. Observed from the special character count is that most phishing URLs have 5 to 10 special characters in the URL, where 7 is the most common. Again this metric is limited by not having data on non-phishing links. From normalising the special character count data it is observed that most phishing links consist of 10 to 20% of special characters.

2 Strengths of the assignment

The report has a nice and clear structure, in which all of the required aspects are discussed. There is a good balance between theory and actual metrics in the report, and the report overall is nice to read. The major strengths we think this report has are as follows:

1. The actors are thought out really well. Not only did you think about the direct actors, e.g. the users and the website owners, but also parties like the government which is mentioned since it is responsible for the digital safety of its citizens.

2. In the ideal metrics section, there is a good connection to the theory/literature of the course, and there is a clear explanation why 100% security is unattainable on a platform, and we should instead focus on perfect balance between cost and benefits/security.

3. Per 'main' ideal metric, you define several 'sub' metrics that help measure/determine the main metric.

4. For every defined metric you clearly explain why you want to measure it, and how it would help platform owners.

5. For every evaluated metric, you (try to) normalise it which ensures that you have a representative view of the metric.

6. The conclusion at the end of the paper, in a single page, nicely summarises what you have done for this assignment, and what the results are.

3 Major issues

1. In the existing metrics section, you talk about a lot of different existing metrics, but almost never refer to literature in which these metrics are discussed. Without these references it cannot be determined if the mentioned existing metrics are feasible, and really used in practice.

2. In the evaluation metrics section, the first three plots have no y-scale. This makes it very hard to determine how serious certain threats/characteristics are.

3. While it is acknowledged in the report, there is no comparison for the special characters in phishing links with the special characters in normal links. While we do have evidence from the literature that phishing links have more special characters, without the distribution of special characters in normal links we cannot make decisions on how much special characters is too much for a URL to be seen as normal.

4. All metrics you evaluate on are more measurements than metrics. From these metrics we cannot deduce how secure we are, we can only get a better idea on how phishing URLs are characterised.

4 Minor issues

1. In the ideal metrics, you mention that "A combination of different types of metrics which cover all controls, vulnerabilities, incidents and (prevented) losses would be desirable to accurately measure and determine security", but for the defined ideal metrics it is not given which metrics fall in which categories, and therefore it is unclear whether the defined ideal metrics cover all categories sufficiently.

2. For the evaluated metrics, one thing that is missing from all metrics is the time component. With the metrics as is, we cannot determine whether we are more/less secure over time.