Measuring cybersecurity - draft

Tycho Teesselink s1560506
Tariq Bontekoe s1451278
Æde Symen Hoekstra s1479679
Ramon Houtsma s1245228

Group 13 - Phishing

Methodology

Steps to follow to solve the assignment:

1. Download the dataset
2. Initial data discovery
    a. Get familiar with the data, what fields do we have?
    b. Open the data in a visualisation program to find interesting relationships
3. Look up relevant information on the problem
    a. What problems does this security issue cause?
    b. What are the consequences of these problems?
4. Look up relevant security metrics from literature
    a. What metrics are there
    b. Which metrics are used in practice
5. Define useful metrics from the dataset
    a. Which found metrics can be defined using the data?
    b. Can we define any other metrics?
6. Evaluate the defined metrics based on the data
    a. How well do we score on the metrics based on the data
    b. Can we define a maturity level based on these metrics?

What we have so far (enumerations will be explained in more detail in final version, and literature will also make sense in the final version as well as the lay-out).

What security issue does the data speak to?

Two ways to look at the security issue of the data:

1. Users click on the phishing links
2. Domains are being misused

The first viewpoint is mainly about social engineering, so getting people to do something they normally would not do willingly, i.e. give up their password or transfer money to a “Nigerian prince”. Phishing attacks usually appear in two forms, either untargeted bulk attacks or targeted spear phishing. These attacks might to lead to great repercussions such as infected computers, installation of ransomware or compromised credentials to ones bank account or digital identification. It is not always that simple to detect these emails. Virus scanners for example will not detect most mails, for phishing mails quite often do not contain viruses.

https://info.phishlabs.com/hubfs/2018%20PTI%20Report/PhishLabs%20Trend%20Report_2018-digital.pdf

According to the report above, there were roughly 16000 known phishing attacks per month in 2017. Contrary to previous years the main target shifted from financial institutions (21%) to email and online services (26%).

Popular attack vectors for phishing are listed below:

- SMS/mobile messages
- Social media
- E-mail
- URL padding (fake login pages)

The second part of the phishing problem is that of domain name abuse. In the current world of the internet each top-level domain is responsible for everything that happens underneath it and can be held responsible by ICANN. This implies that a top-level domains reputation will decrease when the amount of phishing sites within its domain increases. This also implies that these TLDs will try to eradicate phishing sites in their domain, hence it is interesting to look into the abuse within their domain. In an article of Moore and Edelman [ref 7 in paper] observed that a big proportion of URLs containing typos can be found on a relatively small amount of name servers.

https://www.sidnlabs.nl/downloads/publications/speurope2017korczynski.pdf

What would be the ideal metrics for security decision makers?

Expected loss in money, sense of safety, privacy.

Investment in control measures can be substantiated when the expected loss is known. This metric also provides an estimation of how big the phishing problem is within a company/organisation.

Amount of people/companies affected.

This metric shows how many employees or companies are susceptible for phishing attacks. This subsequently shows how important it might be to do awareness campaigns and what effect such a campaign might have.

Ratio of companies/people affected

Type of people affected (location, background). Knowing which type of people are affected can help improve awareness campaigns by specializing the fake phishing. When you can determine which people are easily targeted, malicious people know it too. This metric can help improve the campaigns.

ROI of awareness campaigns.

Knowing the return on investment is a very good metric. It helps on making decisions on how often campaigns need to be done, and whether other investments might be better.

Persistence/occurence of abused domain in TLD.

This metric helps on improving control measures. If certain TLDs are used often, it might be a good idea to handle websites listed under this TLD with caution.

Reuse of attacks under different domain names.

Attacks that are being reused on different domains show that there is still a lack of awareness. People apparently do not notice the body of the attack, only the domain name. This shows that there is still room for improvement in the awareness campaigns.

Type of phishing (banking/social media/…), so trends.

Type of phishing is a good metric, as this helps to understand who are being targeted. These victims can increase their security, and it also helps to improve awareness campaigns by becoming more realistic.

What are the metrics that exist in practice?

We have decided to divide the metrics in practice in four different categories; controls, vulnerabilities, incidents and prevented losses.

Controls

Awareness program click rate. This metric provides an estimate of how security aware employees are. It may also show light on the effectiveness of the awareness programs that have been used in the past, if the click rate is decreasing over time.

Change in behavior of click rate. The change is very important as this shows whether employees are learning. If the overall click rate does not change, but every time a different employee is affected, at least you can infer that the program is working. However, if the same employee is affected each time, the program might not be effective.

Vulnerabilities

Time until detection and response. This metric can be used to determine how vulnerable you are, as well as what impact a successful phishing attack may be. The lower the response time, the better your chances of preventing an attack.

Incidents

Amount of successful attacks over time. This metric provides insight in possible compliance issues due to incidents. It shows whether there are any improvements. This number should definitely be decreasing.

Occurence of Abuse. Difference in amount of abused domain names or top-level domains, in order to see which TLD has the biggest abuse issues.

(Prevented) losses

Phishing emails reported over time. Reporting phishing mails is the most important step. This helps to prevent others from clicking on the phishing email. Therefore the amount of reported phishing mails should be measured.

Persistence of Abuse. This measure will describe how quick TLDs and intermediate registrars are in removing a known abused domain.

Assessing Your Phishing Risks — What Metrics Should You Rely On?
https://cyberready.com/assessing-your-phishing-risks-what-metrics-should-you-rely-on/

Anti-Phishing: Measuring Phishing Awareness Training Effectiveness
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjz4JLty8HdAhXNIlAKHfXxDd4QFjAAegQIChAB&url=https%3A%2F%2Fresources.infosecinstitute.com%2Fcategory%2Fenterprise%2Fphishing%2Fphishing-countermeasures%2Fanti-phishing-measuring-phishing-awareness-training-effectiveness%2F&usg=AOvVaw2tCpzdc8fRBKyVHNjZK7n0

Noticeboard Management briefing on Authentication and phishing metrics
http://www.securitymetametrics.com/61_NB_mgmt_briefing_on_phishing_metrics.pdf

Measure ROI of Phishing Awareness and Education Training
https://www.infosecurity-magazine.com/opinions/measure-phishing-awareness/

Measuring Change in Human Behavior
https://www.rsaconference.com/writable/presentations/file_upload/hum-t07b-security-awareness-metrics_v3.pdf

SLide 14 of presentation by PhishLine called The Fight Against Phishing
http://www.isacantx.org/Presentations/The%20Fight%20Against%20Phishing.pdf

SIDN research on domain names used for phishing websites. (espcially 2.2 Security Metrics)
https://www.sidnlabs.nl/downloads/publications/speurope2017korczynski.pdf

Examining the Impact of Website Take-down on Phishing
https://www.cl.cam.ac.uk/~rnc1/ecrime07.pdf

A definition of the metrics you can design from the dataset

- Number of known active phishing sites at any given timestamp
- Share of known phishing sites per country
- Average lifetime between the reporting of a phishing site and before it is taken down
- Share of known phishing sites still alive since any given time
- Shares of how often different companies are targeted at a given time
- Trends over time of how often companies are targeted
- Which registrars/nameservers are popular for hosting phishing sites?

An evaluation of the metrics you have defined. This should include graphical representations of the metrics (e.g., histograms, scatter plots, time series, bar charts).

We will evaluate the defined metrics in the final report.