Economics of Cyber Security

Block 2 Report

Fiske Schijlen	(4947681)
Alexandros Korpas-Kamarianos	(4907167)
Joël Abrahams	(4443268)
Thanh-Dat Nguyen	(5025664)

{f.h.schijlen, a.korpas-kamarianos,j.s.abrahams, t.m.nguyen-3}@student.tudelft.nl

September 23, 2019

Contents

1 Introduction

Metrics can indicate how security activities contribute to security goals. They can shed light on how effectively the current measures are, and can aid security decision makers in their decision making process for fixing broken processes and improving those present.

We first outline what phishing attacks entail, and what we aim to protect and from whom in section 2. We then outline an idealised set of metrics which would aid security decision makers in section 3, after which we outline our findings of metrics that exist in practice in section 4. These two sections form the framework of 5, where we define metrics using the given data-set. There metrics are used in section 6, where we draw conclusions based upon these metrics.

2 Security Issue

Security is defined by in *The Concept of Security*[1] as "a low probability of damage to acquired values". The author uses a number to questions to define the structure of security, both as a concept and as a policy. Since we are interested about the theoretical definition of the phishing attack as a security issue, it is sufficient to formulate answers for three questions under the examined perspective; *Security for whom*, *Security of which values*, and *Security from what threats*.

2.1 Security for whom?

First, we need to define who needs to be protected, namely "the entity, object, system, unit or the like which is considered to be under threat or which ought to be protected"[2]. In principle, every end-user is a possible victim of a phishing attack. However, this assignment will be focused on organisations as victims. More precisely, an organisation can be defined as any entity/group that consists of people that serves a particular purpose, such as companies or governmental departments.

Organisations, compared to end-users in general, seem to be more interesting as, besides the assets in danger that will be presented on the following subsection, its security is usually more difficult to be achieved. As human factor is really crucial, the fact that organisations consist of a number of end-users increases the difficulty of a good security policy. Hence, even though organisations usually have multiple layers of security, a single human mistake can nullify expensive security measurements and lead to a data breach.

2.2 Security for which values?

Another important step to define our security issue is to describe the values that need to be protected. At first blush, someone could tell that organisations want to protect their assets which are usually important data for their business continuity or private data of their clients. For example, a possible data leakage can cause serious economical damage to a company, either from direct loses (i.e. leakage of confidential business strategies) or from fines (such

as GDPR compliance fines). Furthermore, the reputation of the organisation is disputed, as whenever an organisation is getting hacked, its brand name is hurt.

However, by putting these dangers under the context of phishing attacks, we conclude that the value which need to be protected is the confidentiality of their infrastructures. In other words, the goal of these attacks is to obtain access to sensitive infrastructures of the organisation. This can be done either by intercepting credentials of someone within a certain organisation using a fake login website or by installing a backdoor to an endpoint inside the organisation using a malware delivered with an email.

2.3 Security from what threats?

Last but not least, the definition of the threats is important. Fichtner[2] defines the threat as "an entity or actor deemed not to have certain rights to exert control or access data". In the case of organisations, the phishing attacks are performed usually either from competitors that want to hurt the victim or hackers that want to obtain access to the protected assets of the organisation for their own purposes, i.e. access to users' passwords stored in the organisations' database or sell sensitive data to the black market. Moreover, organisations are often targeted by ethical hackers that discover system weaknesses or hacktivists that have social motives for attacking an organisation.

2.4 Formulation of the Security Issue

According to the aforementioned answers to Baldwin's questions[1], we can formulate the following security issue organisations(who) want to protect the confidentiality of their infrastructures(which values) from third-parties that aim to hit them(what threats), either economically or socially.

3 Ideal Metrics for Security Decision Makers

Metrics can aid security decision makers during their decision making process. Whereas measurements are usually numeric observations, metrics are based on systems and standards, giving them greater context and making them easier to discuss, which is ideal when making decisions. Below, we discuss some metrics which we believe would be ideal for security decision makers, regardless of their actual feasibility. Ideally, a metric is specific, measurable in one way or another, repeatable, and time-dependent. We focus mostly on phishing, though some metrics can be interpreted in a far broader sense.

It should be noted that we attempt to frame these metrics in terms of organisations in general, say for instance government organisations, and not just in terms of corporations specifically. Phishing comes in many forms, such as *clone phishing*, where a fake website tries to mimic a real one to steal login credentials, *phone phishing*, which relies mostly on sms messages, and *DNS-based phishing*, also called pharming, which interferes with the resolution of domain names to redirect users to bogus websites [3].

Security gains from security measures are often intangible in form [4], which, when compounded by the cost of security, can make it difficult to justify the extra cost and potential

burden placed on systems (and users) by implementing said security measures. Therefore, metrics can make it easier to measure the hypothetical gains made by implementing a certain security measures, making it easier to justify its existence. They can also help determine the effectiveness of specific counter-measures, further aiding security decision makers in their decision to either improve counter-measures or replace them with something else perhaps.

3.1 Detection and Prevention Metrics, Measuring the Risk

Detection metrics can be used to determine how eminent an attack might be, or how likely it is that an organisation will be attacked in the (near) future. This is akin to the *likelihood* portion of the *risk* = *likelihood* × *consequence* relation.

Frequent origins of attack | Certain locations, or more specifically servers, are often known to be frequent origins of attack. Emails sent from these locations could be a cause for concern, indicating that the request is more likely to be malicious in intent, meaning extra care should be taken for such requests. Knowing how often emails are sent from such locations can also indicate if a organisation is currently a target of malicious actors, which could perhaps indicate an organisation should be put on alert.

No. of attacks relative to organisation size | It stands to reason that organisations that are larger in size are more attractive to malicious actors, as from their perspective, a successful breach is more likely to reap larger rewards. Accurately knowing the relationship between one's size, and the amount of attacks, can make it easier to assess the risk to an organisation at any given time. This may also make it easier to determine to what extent security measures should scale with organisation size.

Types of attack over time | Following trends in the types of attacks can indicate whether a request made is malicious in intent. Malicious actors are continually thinking of new ways of concealing phishing attacks as legitimate requests. Emails could for instance manually be marked as malicious, which are often identifiable by lack of personal content, invoking a false sense of urgency, and concealing malicious URLs by only using IP addresses, to name a few [5]. Categorising identified threats continually and keeping track of the how common each type is can help identify current, and perhaps future, trends, allowing one to perhaps implement counter-measures more effectively.

Portion of emails that are malicious | What portion of emails that have been flagged as malicious can indicate if the organisation has become a target of attackers, which may indicate that it is important to invest more heavily in counter-measures. This is akin to the *frequent origins of attack* metric, which helps us determine if an organisation is currently the focus of malicious actors.

Employees awareness of phishing attacks | Employee awareness of phishing can reduce the likeliness of a successful attack, as reckless behaviour can often circumvent many safety mechanisms put in place. This is a more difficult metric to accurately asses and measure, as it is difficult to quantify awareness of security issues accurately amongst employees.

Nevertheless, and effort should be made to increase employee awareness, as employee negligence is often a source of security breaches. Somehow keeping track of employee awareness, through use of i.e. survey's, can help an organisation and security decision makers asses policies, and how much should be invested in employee training with regards to (cyber) security.

3.2 Measuring the Consequences

In also useful to measure the potential consequence of a successful attack. This indicates how important it is to actually prevent said attack, indicating how many resources should be poured into preventing it. This is akin to the *consequence* portion of the *risk = likelihood* $\times$ *consequence* relation.

Each metric mentioned below has an implicit *Potential* prefix, as these metrics are based mostly on hypotheticals, ideally extrapolated from current and historic data.

Effect on brand image | A data leak, which may be the result of leaked employee credentials, can be brought to the public's attention, which can results in (lasting) damage to a brand's image and reputation. A damage to reputation could, in the case of a corporation, lead to decreased sales, potentially causing lasting financial damage long after the attack itself. Therefore, a method of measuring the effect on a brand's image would give a more accurate image of the actual cost of a security breach.

Amount of users lost | Due to several security breaches over the years, it is easier to gauge how strongly users react to data breaches. This ties in closely with *Effect on brand image*. This is mostly relevant to corporations.

Legal ramifications | Depending on the type of data leaked, there can be legal ramifications for leaking said data. Whereas leaking a set of usernames to the public might damage reputation as it could give notice to perhaps sloppy security practices, leaking credit card or bank information might lead to lawsuits to compensate for potential financial damages inflicted upon users of a system.

Amount of money lost | The amount of money that may be lost after a breach can directly indicate how much money should be invested in preventing phishing attacks. For instance, core business assets (in the digital domain) might cause larger financial fallout for an organisation than minor, non-safety critical information. While it might be difficult to assign a concrete value to information, knowing the relative value can help determine what should be protected.

This is also relevant when considering legal ramifications, greater care should be taken to protect data which poses a greater legal risk if leaked.

4 Existing metrics in practice

The metrics defined in section 3 provide an idealised picture, reflecting what security decision makers wish to know. However, such idealised theoretical metrics can be difficult to quantify and measure accurately. It is, however, possible to build upon such idealised metrics to produce metrics which attempt to fulfil these wishes to one extent or another, meaning that these idealised metrics function as a guide of sorts when constructing metrics which may be used in practice. Below we outline some of these more practical metrics.

click-through rate: This metric measures the number of times that users fall for phishing emails.The rate may be for a specific phishing email as well. Most of the time, this metric is used for phishing awareness programs where users face simulated phishing emails. It allows tracking whether the users become more conscious of this attack and hence the program's effectiveness. There are flaws to this metric. The metric does not consider how elaborate the phishing email is and how well designed it is. As such, the click-through rate may increase leading to misinterpretation of the failure of the program. Another flaw is when there is no steady user base. New users will likely lead to a rise in the click-through rate. [6, 7, 8]

click-through rate per user This is a variant of the above mentioned metric that focuses on one single individual. It reflects the user's resilience to phishing emails. This metric allows to identify users that are prone to phishing attacks. [7, 9]

average detection/report time This metric measures the time until a phishing email is reported by a user or detected by the IT security team. It allows to track the response time to phishing emails. [10]

number of reported phishing emails reported This metric measures how many phishing emails are reported. This is particularly useful in phishing awareness programs where the number of simulated phishing emails are known. It reflects the effectiveness of such a program. [10, 11]

targeted brands by phishing attacks This metric tracks what brands are targeted the most by phishing attacks. [11]

number of phishing domains The metric measures the number of reported and detected phishing domains. This can be used to monitor phishing activities over time. The metric can be narrowed down further by measuring phishing domains per Top-Level Domains (TLD).[12, 11]

average up-time of phishing site Tracking the average up-time of detected phishing site by TLDs shows how effective phishing detection is. The shorter the average up-time, the faster phishing sites are being detected and shut down. [12]

average traffic TLDs can track traffic of known phishing sites to estimate their effec-

tiveness. High traffic in a phishing site could indicate that a lot of users are falling for the phishing attack and hence are accessing the phishing site. [12]

phishing domains per ISP Phishing domains are not equally distributed in ISPs. Some host a larger number than the average. Such a metric can be a good tool to identify traffic that should be more closely monitored or even blacklisted. [12, 11]

most targeted industry sectors This metric measures the distribution of phishing attacks in different industry sectors. It allows to track how much the targets of phishing attacks are shifting over time. [13]

percentage of phishing attacks hosted on HTTPS The number of phishing domain hosted on HTTPS is tracked. Modern browsers usually warn when SSL is not used which can be a first indicator of a phishing attempt. The metric tracks how many attackers manage to get a certificate for their phishing domain.[13]

5 Metrics that are applicable to the dataset

5.1 Relevant data attributes

In practice, you can only make use of metrics which the dataset allows you to produce. Our phishing dataset contains several attributes that can be used for this purpose:

- `firsttime` and `lasttime`, which can be used to determine the threat's uptime and whether it has been recently active;
- `virusname`, which actually refers to the entity or organisation as which the phisher is acting and can be used to find out which organisations are most commonly used in phishing procedures;
- `recent`, which shows whether the used phishing link is still up or not;
- `ip`, `domain`, and `email`, which can be used to determine commonly occurring phishing sources;
- `country`, which can be used to determine which countries have the most active phishing attempts.

5.2 Practical metrics based on the dataset

Our dataset describes several properties of phishing sources, but it does not contain source- or target-specific information that could determine the effectiveness of the phishing attempts. Specifically, we have no information regarding click rates, detection times, and user reports, which rules out the practical metrics from [6], [7], [9], and [8].

We can, however, use the `firsttime` and `lasttime` attributes to estimate the mean uptime per TLD, country, IP, or domain. This can then be used to determine the threat level of certain phishing threats. Specifically, we can use them to implement the activity of phishing sources in certain time frames as is often done in practice[13], so an organisation would be able to measure the effectiveness of their security controls.

Finally, we can estimate proportions of phishing sites that contain certain patterns in their domain names. For example, we could estimate the proportion of phishing sources that use HTTPS, which we could use to more accurrately identify if a given source is actually a phishing source.

5.3 Metrics that can be defined using the dataset

Using our dataset, we can measure phishing sources' threat levels by comparing their uptime, recent activity, and frequency. We can then use these threat levels to gather phishing domains and IP addresses to blacklist. We believe this approach to be effective because blacklisting is one of the most popular anti-phishing solutions.[14]

With threat level estimation in mind, we define the following metrics *for each occurring phishing source in the dataset*, where we treat a row's IP address as a phishing source:

- Source IP occurrence frequency;
- Source country occurrence frequency.

Furthermore, we define the following metrics to evaluate the effectiveness of implemented security measures, as well as the general activity of phishing sources in certain time frames:

- The average phishing source uptime per month;
- Percentage of phishing attacks hosted with HTTPS.

6 Evaluation of Metrics

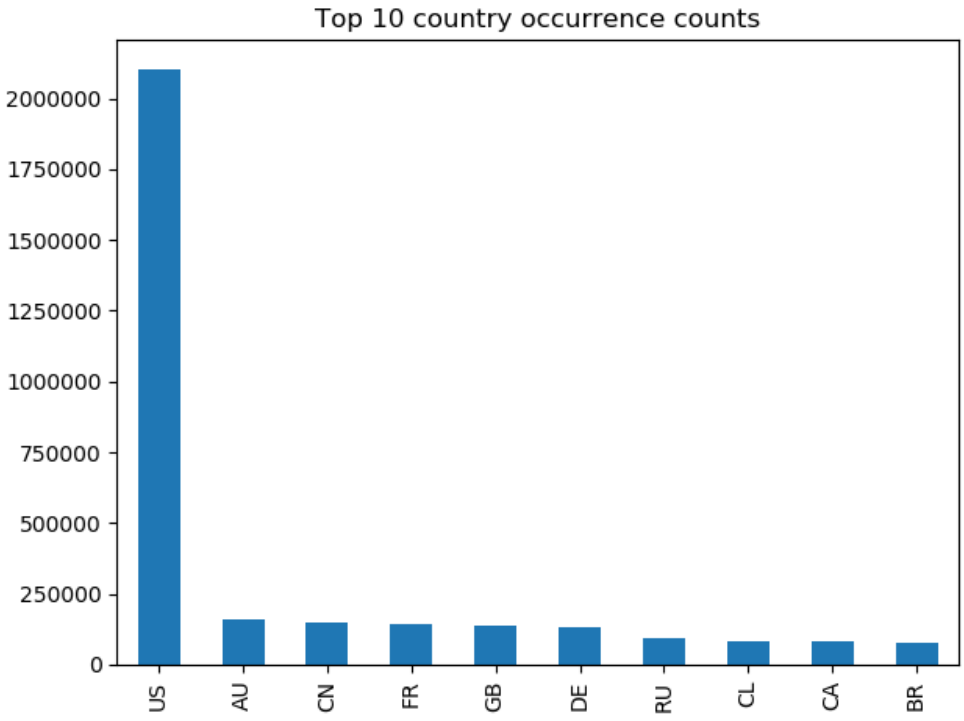

Figure 1

The data from figure 1 is not very helpful for security decision makers. The goal here is to find out which countries produce more phishing traffic, so we can figure out which traffic we should monitor more. However, these occurrence counts also heavily reflect the total amount of traffic produced by each country. This means the data is useless to decision makers unless it is somehow normalised, e.g. by comparing these counts to the total amount of traffic received by a specific organisation from a certain country. We cannot do so with our data set, because it only contains phishing data.

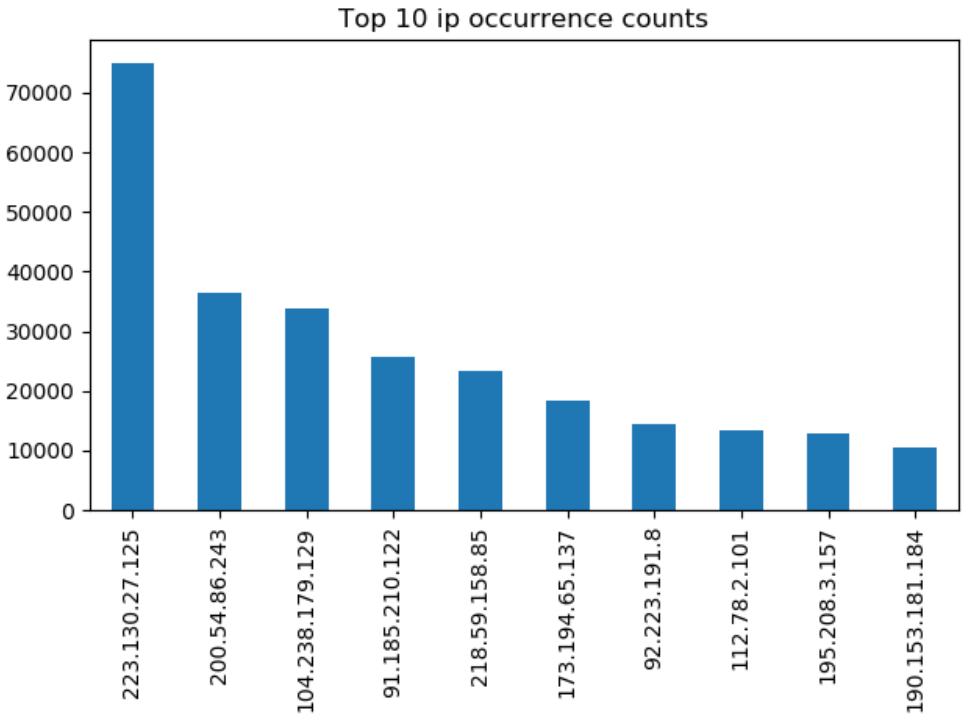

Figure 2

Figure 2 shows that plenty of IP addresses occur way more than others. Security decision makers can use this information to define a threshold for IP occurrence counts and blacklist those of which their count exceeds the threshold.

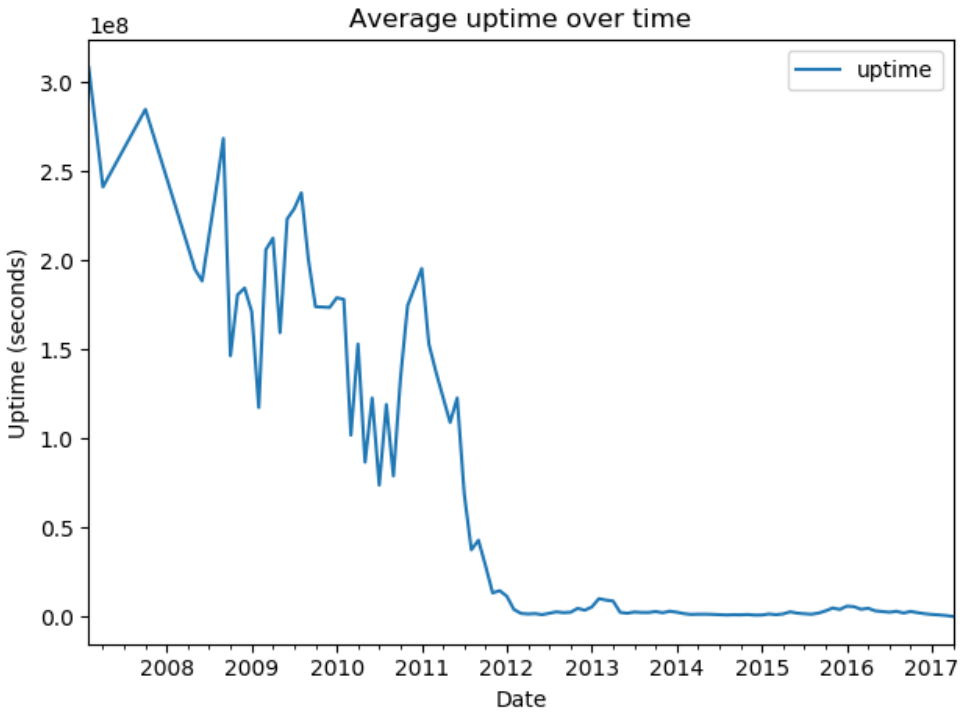

Figure 3

Figure 3 shows the average up-time per month of phishing domains that were started that same month. It can be used to measure the effectiveness of phishing domain detection. If the up time decreases, one can conclude that phishing domains are being detected and shut down faster. The graph shows that the average up-time of a phishing domain decreased until around 2012, and then stayed relatively constant. Some assumptions were made to plot the graph. For domains that are still active according the data, we assume that the domain was still active when the data was retrieved. We also assume that the date of retrieval is the most recent date that can be found in the data set. The limitations of the metric is that it does not include undetected phishing domains. The data used might not model the real situation correctly. For example, phishing domains are getting better at avoiding detection. Thus, the graph would show a decrease indicating an improvement in the detection.

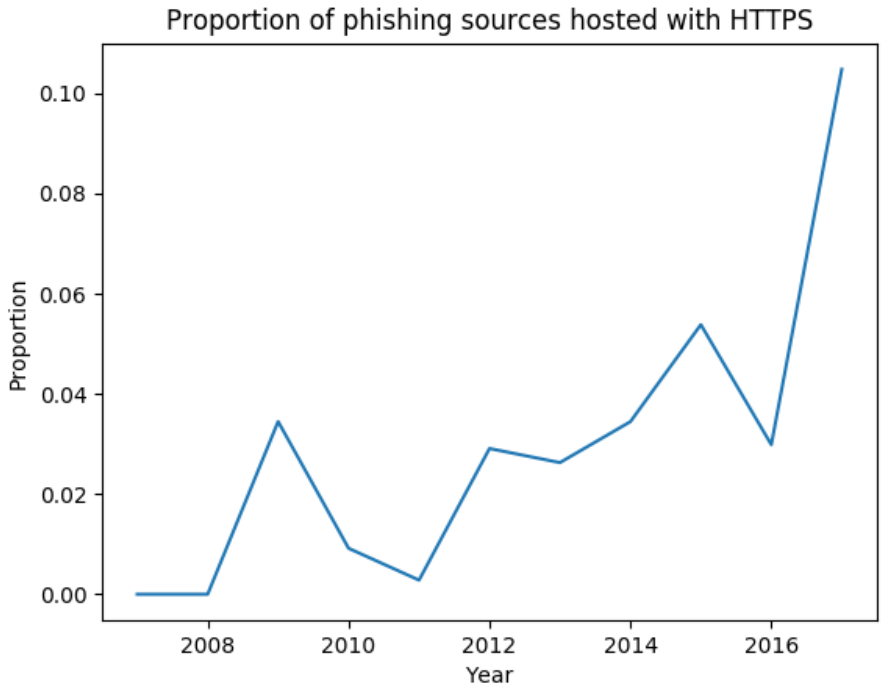

Figure 4

The yearly proportions of HTTPS-hosted phishing sources shown in figure 4 neatly correlate with those from [13]. The trend implies that phishers are focusing more on hosting their domain with HTTPS each year, likely so that their victims will trust the URL more easily. Decision makers can use this data to inform their employees that HTTPS-hosted URLs are not necessarily safe, despite what they may have assumed in the past.

However, our data set only contains known phishing sources. The undetected sources may contain significantly more HTTPS-hosted URLs, which means our data may be somewhat inaccurate.

7 Conclusion

Using our data set, we have attempted to define metrics that could help security decision makers make decisions to alleviate the impact that phishing attacks have on their organisation.

Ideally, we would like to have had a data set that describes the frequency of attacks and the success they had. However, most public data sets do not provide such convenient attributes.

Our data set allowed us to define the following useful metrics:

- Source country occurrence frequency;
- The average phishing source uptime per month;

- Percentage of phishing attacks hosted with HTTPS.

The latter 2 metrics were heavily inspired by the practical metrics mentioned in section 4. Our evaluation of these metrics has shown that they can be used to determine a significant amount of phishing IPs to blacklist, estimate the effectiveness of phishing attacks over time, and determine that HTTPS-hosted phishing URLs are becoming more prevalent. Using these metrics, decision makers could decide to update their firewall blacklist, estimate the effectiveness of their security controls, and decide whether or not employees require more awareness training with regards to phishing attacks.

References

[1] D. A. Baldwin, "The concept of security," *Review of International Studies*, vol. 23, pp. 5 – 26, 1997.

[2] L. Fichtner, "What kind of cyber security? theorising cyber security and mapping approaches," *Internet Policy Review*, vol. 7, no. 2, 2018. [Online]. Available: https://policyreview.info/articles/analysis/what-kind-cyber-security-theorising-cyber-security-and-mapping-approaches

[3] M. N. Banu and S. M. Banu, "A comprehensive study of phishing attacks," *International Journal of Computer Science and Information Technologies*, vol. 4, no. 6, pp. 783–786, 2013.

[4] R. West, "The psychology of security," *Communications of the ACM*, vol. 51, no. 4, p. 34, 2008.

[5] M. Chandrasekaran, K. Narayanan, and S. Upadhyaya, "Phishing email detection based on structural properties," in *NYS cyber security conference*, vol. 3. Albany, New York, 2006.

[6] H. Siadati, S. Palka, A. Siegel, and D. McCoy, "Measuring the effectiveness of embedded phishing exercises," 2017.

[7] CybeReady. Assessing your phishing risks — what metrics should you rely on?n. [Online]. Available: https://cybeready.com/assessing-your-phishing-risks-what-metrics-should-you-rely-on

[8] K. R. (2018) The trouble with phishing. [Online]. Available: https://www.ncsc.gov.uk/blog-post/trouble-phishing

[9] CybeReady. Is click rate the right way to measure phishing simulation training? [Online]. Available: https://cybeready.com/phishing-simulation-training-right-metrics

[10] R. Fahey. Anti-phishing: Measuring phishing awareness training effectiveness. [Online]. Available: https://resources.infosecinstitute.com/category/enterprise/phishing/phishing-countermeasures/anti-phishing-measuring-phishing-awareness-training-effectiveness/

[11] AWPG. (2016) Global phishing survey. [Online]. Available: https://apwg.org/globalphishingsurvey/

[12] M. Korczynski, S. Tajalizadehkhoob, A. Noroozian, M. Wullinkz, C. Hesselmanz, and M. van Eeten, "Reputation metrics design to improve intermediary incentives for security of tlds," 2017.

[13] AWPG. (2019) Phishing activity trends report. [Online]. Available: https://docs.apwg.org/reports/apwg_trends_report_q1_2019.pdf

[14] N. Abdelhamid, A. Ayesh, and F. Thabtah, "Phishing detection based associative classification data mining," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5948 – 5959, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417414001481