Don’t Forget the Human: a Crowdsourced Approach to Automate Response and Containment Against Spear Phishing Attacks

Pavlo Burda
Eindhoven University of Technology
p.burda@tue.nl

Luca Allodi
Eindhoven University of Technology
l.allodi@tue.nl

Nicola Zannone
Eindhoven University of Technology
n.zannone@tue.nl

***Abstract*—Organizations are increasingly facing sophisticated social engineering attacks that exploit human vulnerabilities and overcome commonly available countermeasures. Spear-phishing campaigns are becoming the most prevalent attack and source of compromise for most organizations. We argue that existing prevention and detection countermeasures are fundamentally ineffective against this class of attacks. In this work, we propose a novel approach to address the limitations of existing countermeasures. Our proposition is a new course of action to exploit human detection capabilities as a basis of automated response strategies. Preliminary results unveil users’ mental models for phishing detection and reporting as a way to improve the phishing reporting process altogether. A real word case study is provided to promote the feasibility of our proposal.**

***Index Terms*—spear phishing, counter-measures, human measurement, containment**

1. Introduction

The cyber-threat landscape transition from purely technical (e.g., automated malware delivery, remote vulnerability exploitation) to socio-technical exploitation methods such as phishing attacks brings new challenges for defending the ever more integrated system infrastructure. On this wave, a major attack vector is the email, which is used by social engineers to target the general population, organizations and even individuals. Phishing techniques provide the advantage that the ‘human vulnerabilities’ they attack cannot be easily fixed, and are approximately the same across targets. For this reason, phishing attacks are evolving rapidly into new, more sophisticated attack scenarios from the usual ‘*your email account is expired, click here to renew your password*’ and to overcome available countermeasures. Being this a technologically cheap yet powerful exploitation technique, it has become the preferred method employed by attackers to compromise systems and exfiltrate data from individuals as well as large targeted organizations [1].

The latter in particular represents an increasingly worrisome trend of sophisticated, highly-targeted social engineering attacks [1] against which common countermeasures aiming at ‘general’ phishing are almost hopeless [2]. These attacks are ‘tailored’ against specific organizations or groups of people, and differ significantly from generalist attacks. For example, by means of multiple iterations and reconnaissance, an attacker can tailor social engineering artifacts that are extremely effective on their targets [3]. Cognitive attacks aimed at persuading victims in executing an action are diluted in multiple interactions exploiting the communication methods and language the organization is used to, making them hard if not impossible to detect by traditional means.

Following the NIST *protect*, *detect* and *respond* framework, state of the art counter-measures integrate training of employees, advanced software and security operation centers. However, existing countermeasures are lagging behind the expansion of sophisticated phishing attacks, first of all, spear-phishing [4]. Attack features like content and links are extremely variable, hindering the majority of detection attempts or generating too many false positives [5]. Further, the resemblance of these attacks to regular communication make training and awareness campaigns largely ineffective to ‘immunize’ a large fraction of the victim pool [6]; anomalies in the communication still exist (e.g., unusual references to internal processes in an organization), but these are hard to formalize and cannot be captured automatically by a single technological solution. Therefore, organizations often rely on response teams, like SOCs and CERTs, as the last line of defense. However, current containment procedures based on after-the-fact analyses are too slow to match the high velocity at which spear-phishing campaigns are known to affect their targets [7].

In this position paper, we propose a way ahead to respond and contain advanced spear-phishing attacks in organizational settings. Our proposition aims at merging the natural ‘immunity’ of (some) human subjects in an organization (e.g., senior employees with a deep knowledge of the ‘normal’ processes within the organization and a natural ability to detect ‘anomalies’) with automated response procedures to mitigate and contain the attack against the organization as a whole. At the core of the proposed solution is a more efficient *phishing reporting process* based on cognitive mental models of individuals better predisposed to detect complex attacks. The improved reporting process will allow to speed up the containment of an attack, thus lowering the number of victims.

The remainder of the paper is structured as follows. The next section discusses previous work on phishing counter-measures and highlights the identified research gaps. Section 3 provides a description of the proposed solution, a motivating example, and our research plan. Section 4 presents preliminary results. Finally, Section 5 concludes the paper.

DOI 10.1109/EuroS&PW51379.2020.00068

2. Related Work

The phishing phenomenon has been extensively studied in the literature, with a particular focus on the design of technical countermeasures based on blacklisting and machine learning classification [8], [9]. Despite these efforts, no definitive solution to phishing has been found yet. The underlying issue is the lack of a deep understanding of the complex socio-technical problem entailed and exploited by phishing [10].

Phishing campaigns can take various forms, depending on the value of the target and required resources. The most common form of phishing attacks aims at large numbers of victims and often contain cognitive exploits to persuade targets into making wrong decisions in a 'hit-or-miss' fashion. On the other side of the spectrum, we can find spear-phishing attacks, which are typically devised to target single individuals or small groups of individuals and are characterized by iterative reconnaissance and attack engineering phases and by the sophistication of the artifacts employed in the attack [2].

While the mean susceptibility rate to phishing attacks across various experiments and measurements is 21% [8], spear-phishing exhibits more impressive numbers: 80% cadets in a military academy were successfully spear phished in a training exercise [11]; Kumaraguru et al. successfully spear phished around 50% of the subjects in their experiment [12]. In their *context aware* phishing campaign against Indiana University students, Jagatic et al. obtained a success rate of 70% [13]; Caputo et al. report a spear-phishing susceptibility rate of 60% in their first trial [14]; Burns et al. obtain a click rate of 70% [6].

The high susceptibility rates achieved by spear-phishing indicate that current countermeasures might not be well-suited against this type of attacks. Next, we discuss the various methods to counter phishing attacks and their effectiveness against spear phishing, starting from solutions employed to reduce phishing susceptibility of potential targets to response strategies like infrastructure take-downs.

Prevention. The ideal remedy to spear-phishing and, in general, to phishing is to make potential targets immune to the attack altogether. Preventive measures typically encompass the training of users to recognize specific attack features and rising their awareness of the threat. Prior studies show that training has a significant effect in reducing generic phishing susceptibility, albeit leaving margins of untreatable portions of subjects around 10-15%, even with repeated training [15], [16]. The same effects, however, might not be achieved against spear-phishing. Kumaraguru et al. performed a controlled field experiment to test the effectiveness of training tailored to spear-phishing, showing no significant differences between generic and spear-phishing training effects [15]. Caputo et al. report no effects of training (and awareness) at all when conducting a spear-phishing attack in their experiment [14]. Burns et al. report a marginally significant effect of training tailored to the detection of spear-phishing attacks, reducing phishing susceptibility rate from 70% to 54% after five weeks of training [6]. Other works show that training effectiveness decreases over time [17], [18]. Even if some reduction can be achieved, the underlying problem of the training and awareness campaign lies in the fact that spear-phishing can take very different forms, making the attack difficult to be recognized by users, and requires much less victims than generic phishing to achieve the desired objectives [14]. As some users will still remain vulnerable, training and, in general, preventive measures alone may not minimize the attack surface enough to neutralize or effectively contain spear-phishing attacks.

Detection. The most popular approach to the detection of phishing attacks is *artifact filtering* in an anti-spam fashion, including emails, URLs and attachments [4], [5], [19]–[22]. These countermeasures have been implemented using numerous methods, like data mining, machine learning, heuristics and white/black listing [10]. Solutions based on machine learning techniques might be affected by a large number of false positives and require continuous retraining [5]; they also are not generalizable across domains [23]. A few studies show how these solutions can be bypassed, for instance by legitimizing the sender (via multiple iterations) to appear less "anomalous" to an anomaly detection system [5] or by taking over a legitimate account, e.g. one of the target's secondary accounts or one of her associates [24]. Another body of research focuses on the examination of phishing sites and server characteristics and relies on blacklisting. Some of these works leverage crowdsourcing [25], [26] and reputation systems [27] to improve accuracy and speed. While these solutions have proven to be suitable against general phishing and known threats, they face significant limitations against spear-phishing, as blacklists do not generalize well to unknown [28]. The fundamental drawback of automatic detection techniques against spear-phishing is the unforeseen nature of attack characteristics and artifacts, like pretexts and links [2], [3]. Such artifacts are meticulously crafted to fit targets' context, like demographics, work and previous social interactions [3] and to fly under-the-radar by employing legitimate-but-compromised or vanilla websites and targeting a small numbers of recipients [5]. Detection can also be accomplished by security analysts, who often are superior to automatic tools. However, this solution requires that artifacts of interest are first reported to security analysts by the targets themselves. Thus, these countermeasures leave advanced attacks to remain undetected or to be detected too late when the attack may have already propagated.

Response. Response strategies are typically employed by an organization's IT department and security operation center (SOC) to mitigate the damage of an attack when it occurs. Response teams are aided by both detection techniques outlined above and notifications from users that detect and report the attacks [7]. A response can be initiated immediately after detection (e.g., employees notifications) or later after the attack effects have manifested (e.g., a data leak was identified) [1]. In the former case, incident reporting can alert interested parties (e.g., subsidiaries or clients) of the incoming threat [2]. The remediation procedure can combine attack interception by blocking traffic or investigating rogue domains, although the attacks is typically characterized by a short duration, hindering such attempts [3]. In particular, these containment procedures are sensitive to time delays [28], especially in case of unknown attacks [7]. Prior studies have shown that the response time often does not mach the velocity of spear-phishing attacks where the compromise and exfiltration timelines are skewed towards minutes and hours, while discovery and containment are in the order of hours and days [1]. For instance, Jagatic et al. observed a 50% success rate after 6 hours from the launch of the attack [13], whereas the same rate was achieved within only 2 hours in [12]. To the best of our knowledge, these are the only studies that investigated the velocity of attack propagation in the context of phishing.

Research Gap

Existing countermeasures aim to reduce the victimization rate. However, while they have proven some effectiveness against generic phishing, they are inadequate against spear-phishing. Training and threat awareness are unable to make subjects immune to sophisticated attacks, leaving a large fraction vulnerable to them. Similarly, detection techniques are not able to cope with the large variance in spear-phishing attacks, including chosen pretexts, single vs. multi-stage attacks, and the 'dilution' of specific attack signatures across multiple communications or phases of the attack. Anomalies in the communication processes and protocols characterizing a specific organization may represent an unexplored venue for research, but these processes and protocols are hard to formalize; as a result, a general anomaly-detection solution for spear-phishing applicable to any organizational settings is not on the horizon. The fundamental problem is that the specific characteristics of spear-phishing attacks (multistage processes, tailored artifacts, yet-to-be-seen malicious URIs, etc.), make current defensive techniques inadequate and ineffective.

The consequences are well signaled by industry reports which point at phishing attacks to be the most prevalent attack and source of compromise for most organizations [1]. Therefore, new approaches are necessary to cope with spear-phishing attacks for which both prevention and detection are fundamentally unsuited as the prevalent defensive barrier.

3. Proposed Solution

Due to the foundational differences between phishing and spear-phishing, prevention and detection techniques may be grossly inadequate to tackle the problem. However, we believe there is an important gap in the *response* phase that could provide large benefits to organizational security: human reporting is an untapped resource that could provide readily available risk indicators for suspicious emails, and lead to fast attack response and containment. This requires increasing the quality of phishing reports and automating a risk-based containment phase to promptly react to a reported attack. Preliminary evidence (see Sec. 4) suggests that some users are naturally predisposed to identify anomalies between the communication processes employed by spear-phishing attacks and the 'normal' ones employed by an organization. However, only a few users typically report phishing emails, and the rationale and incentives behind this are still unexplored in the scientific literature. Once deconstructed, the *mental models* behind phishing reporting could be employed to increase reporting incidence, speed, and to build reputation-based methods (like in [27]) to assign risk-scores to specific (likely) attacks. Moreover, the few users that report suspicious emails do this *immediately* when they receive them, providing a timely information source to employ for response. Yet, this is currently untapped. Based on user reports, a portfolio of automated response strategies can be triggered (e.g., issuing warnings to other users, automated URI blacklisting, AV signatures generation, centralized filters, etc.) to protect the large fraction of users that have not *yet* fallen for the attack, but that most likely would if no response is put in place.

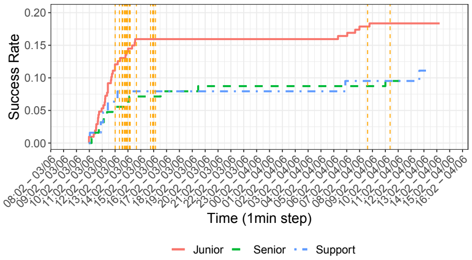


Figure 1: Success rate over time per user category. The vertical lines indicate the time our emails were reported.

Motivating example. Previous finding in the literature [12], [13] already showed that (spear-)phishing attacks trigger victim responses very quickly. However, it is unclear what is the potential of reporting mechanisms to provide timely information on the attack. To provide a first evaluation of this, we look at a tailored phishing campaign we ran against our university for internal measurement purposes [29], in collaboration with the security department of the university. The campaign pretext asked users to participate in an HR process to collect vacation hours, a process that is *not* employed at our university. Figure 1 reports the rate of users falling for the attack (i.e., that would have submitted their credentials if this was a real attack) by user category. Notice that Junior employees (PhD students and postdocs) are those that fall for the attack the quickest and by the largest fraction. By contrast, senior scientific staff and support staff are much less vulnerable overall, further supporting the intuition that expertise on internal processes may be a decisive factor in the successful distinction of a spear-phishing attack of this type. Regarding the velocity of the attack, approximately 75% of employees that fell for the attack did so in the first four hours since the start of the campaign. Interestingly, 23 employees detected and reported the attack to the IT department, many of which when the campaign was at its peak (vertical lines in Figure 1). Our intuition is that a containment action during the (first few) incoming reports can eventually reduce the victimization rate by 25 to 40 percent by automatically blocking the attack.

Research Plan

Our plan to make the security process more efficient is divided in two parts. In the first part, we aim to leverage the already present human detection capabilities of 'phishing champions' to improve the reporting process. In the second part, we intend to use the improved reporting process to automatically launch a containment measure for attacks that could not otherwise be addressed in time.

Methods to improve the reporting process. From an organizational point of view, companies can benefit from employees that notify the IT departments in case of abuse.

The efficiency of this reporting process, however, depends on the number and the quality of such notifications. We hypothesize that, among the employees of an organization, there are some that are particularly good at detecting phishing, further down referred to as ‘phishing champions’. However, not all of those are keen to report suspicious emails. We are interested in fostering reporting only from champions to keep the noise/signal ratio to favorable levels.

We plan to devise methods able to identify phishing champions, and to do so, we first need to understand what are their characteristics. We look for characteristics that may correlate with higher detection skills (e.g., experience) and reporting eagerness (e.g., sense of responsibility). We can collect this qualitative data by means of structured interviews of an organization’s employees that have reported phishing attacks in the past and potentially those who have not reported but detected them. By employing coding techniques to analyze the interviews’ answers, we can extract actionable topics and reconstruct the *mental models* users follow when deciding whether to report a phishing email. Mental models go beyond simple schemata of highly regular and routine situation (like trivial phishing) and can better represent new situations through the use of generic knowledge of space, time, causality, and human intentionality [30].

Based on these mental models, we can design a diagnostic test to systematically identify ‘champions’, including those that are not keen or aware of the reporting process, and develop risk-based metrics to evaluate the uncertainty around the report. These metrics can be based both on past reporting activities of the employee, as well as specific characteristics of the reported artifact.

Methods for automated response to spear-phishing attacks. Having a reliable reporting process and a defined risk metric does not address the time issue per-se. It remains crucial to operationalize the risk metrics to anticipate the containment phase as soon as possible after the first user notification, thus limiting the attack propagation. The preliminary results shown in Figure 1 suggest that notifications may indeed arrive ‘soon enough’ to enable this strategy. To operationalize this idea, we plan to investigate an automatic response procedure that can be initiated when sufficiently many high-risk reports are collected.

One of the possible challenges that can arise, is the decision of how many notifications are necessary to trigger a response, and how to assign weights to different phishing champions. If a too strict threshold is chosen, it may generate a false positive; if too loose, more time will be necessary to collect more reports and, by the time, the window of opportunity can shrink considerably. A testing phase would be necessary to establish a balanced trade off. However, fine-tuning is a significant challenge as well, since the threat we are dealing with is of an infrequent kind. Simulations in the form of realistic phishing campaigns can be a viable option to address this issue.

4. Preliminary Results

As a first effort towards the identification of phishing champions, we interviewed the employees that reported our phishing attempts to the IT department during the phishing campaign reported in Figure 1. Specifically, we were able to interview 12 out of the 23 employees that reported our phishing email. Following a one-page interview guide [31], we first asked high level questions on detection and reporting. Then, we invited the interviewee to retrieve and read the e-mail they received (if needed, we provided a printed copy) and asked detailed questions with the specimen at hand. The interviews investigated how does the interviewee:

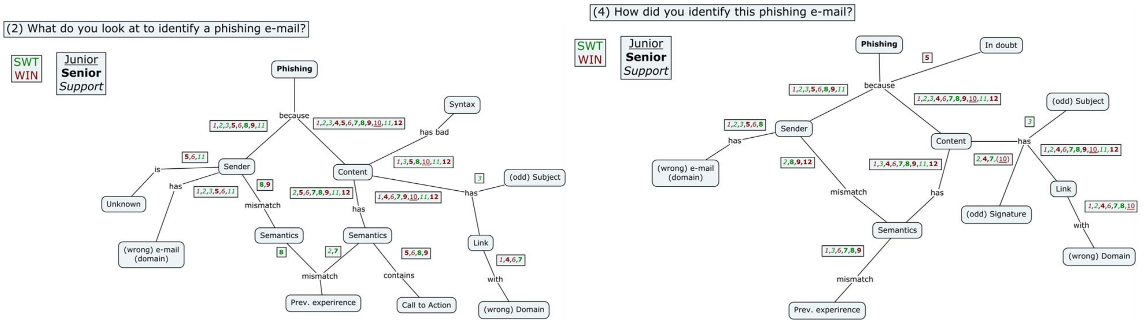
1) detect phishing emails in general and the specific email they received
2) decide to report phishing emails in general and decided to report the specific email

The semi-structured interviews were recorded and transcribed. The interviewees’ answers were coded using a card-sorting technique to derive mental models reflecting the decision process of the respondents. The more similar users’ thoughts are between the general case and the specific attack, the more ‘mature’ the mental model can be considered to be, as it characterizes users that can abstract their reasoning away from examples without loss of information. By contrast, mental models that are much more detailed when example-driven than in the general case suggest a less mature rationale whereby users cannot abstract away from the example.

Results

Figure 2 presents the results concerning the detection of phishing emails in general (left) and the detection of the phishing email sent within our campaign (right). We include subjects’ characteristics like seniority and department they work at. Arcs are labeled with numbers identifying the interviewees which followed that arc in their reasoning. For example, in Figure 2b, ID 7 is a senior employee whose detection method can be reconstructed following the arcs labeled with ID 7, starting from the root: 1) the content’s semantics does not match her previous experience, 2) an unfamiliar signature is present, and 3) there is a link with a wrong domain name. We can observe that, for detection, the mental model derived from the concrete example (Figure 2b) largely matches the mental model derived from the general case (Figure 2a). In contrast, there is not a clear overlap between Figures 3a and 3b, suggesting a less developed mental model for reporting.

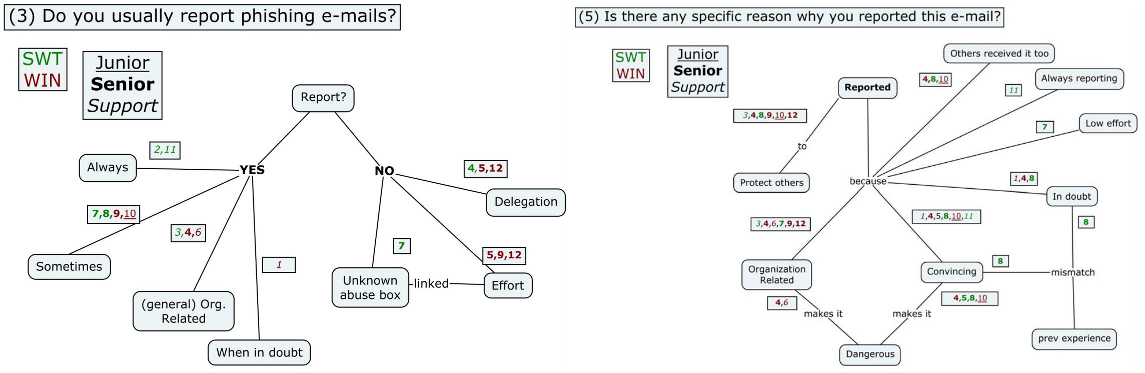
How do our interviewees detect phishing attempts. In the interview, we first asked users how do they detect *general* phishing attempts, and why did they detect our specific phishing email. Answers to the first question prevalently refer checking if the content’s syntax is correct, if the semantics ‘makes sense’ to the user and if the sender’s email domain is correct. When prompted with the email from our campaign, the reasons why our respondents detected the phishing email as such largely overlap with the answer they gave us in the general case. For example, Respondent 3 states “*It’s a follow up? Strange request. Why is there a link? Strange email...*”, highlighting the ‘anomalous’ nature of the email w.r.t. what she is used to receive from the department. Similarly, Respondent 8 answered: “*It’s a bit weird for UNI to ask me my holiday hours, I already have them on the [HR’s portal]. Also the domain is not good [..]. It’s asking for specific action*



(a) General case

(b) Concrete case

Figure 2: Mental model of phishing detection



(a) General case

(b) Concrete case

Figure 3: Mental model of phishing reporting

which does not apply to me.”, further highlighting the email inconsistencies also remarked by many for the detection of a ‘general’ phishing email. On this same line, Respondent 9 answers: “*The sender does not match the topic semantics, it’s like you have a painting from Rembrandt and suddenly you have an iPhone there*”.

In all, we find our respondents were very consistent in their rationale to classify a phishing email as such, whether this is a ‘general’ or hypothetical phishing email, or a concrete example with which they are already familiar.

Why do our interviewees decide to report phishing emails. When asked if they usually report phishing attempts, the majority of respondents answered that they do not report phishing on their personal (email) accounts, and reporting at work happens more as an exception. As for the question why they do report phishing emails, answers are very general and a clear rationale does not emerge. For example, Respondent 1 answered: “*[I report when] I’m in doubt it could be a legit email*“, highlighting uncertainty as a key element in their decision to act on the attack. On a similar line, Respondent 6 states: “*I report if I know the sender, e.g., my bank or my organization*”, suggesting that only emails that are ‘relevant’ to the user’s context will be reported by this user. However, a clear-cut reason to discern between ‘general’ phishing emails that our respondents will report, and those they won’t did not emerge.

By contrast, the reasons to report our specific phishing email appear to be much more structured and detailed, and include reasons relating to safeguarding less-aware colleagues and the perceived sophistication of the attack. For instance, Respondent 3 states she reported the email because “*it pretends to be from UNI, to protect others*”; similarly, Respondent 6 says he reported the email “*Because it’s posing as UNI, my organization should know about it*”. In sharp contrast with the detection case, the respondent’s mental models for phishing reporting appear to be much less developed, structured, and consistent, suggesting a strong imbalance in user prowess between detection and reporting activities.

Discussion

Mental models showed the respondents’ inability to generalize the rationale for notifying a suspicious email, thus providing insights on where improvements of security processes can be made. For example, answers highlight that the reporting procedure is ill-perceived in terms of effort and liability (“Effort” and “Delegation”[1] in Figure 3a). Such models can also shed light on the underlying factors for reporting, like a higher sense of responsibility and threat awareness (“Protect others” and “Dangerous” in Figure 3b).

The results, however, may be influenced by the specific type of organization where the study was carried out. Other domains, like financial or industrial, may lead to different

[1] By “Delegation”, the respondent assumed it is someone else’s duty to deal with security incidents.

outcomes both in terms of reporting rates and reasons to report. More studies are needed to generalize our results.

From our preliminary evaluation, a more thorough and rigorous investigation could shed additional light on the following research question:

> *What mental models do users follow when deciding to report a phishing attack, and can those models be improved to better support an organizastion's security processes?*

Future work could tackle new research in this direction by evaluating the training and reporting problem, for example by investigating whether an efficient phishing *reporting* process can aid the protection of users that fail to detect the phishing email as such.

5. Conclusion

In this position paper, we argued the urgency for new paradigms to counter spear-phishing attacks. In particular, we proposed a new course of action to address the limitations of existing countermeasures against this class of attacks by exploiting human detection capabilities as the basis for automated response procedures. We are guided by the intuition that a sufficiently high proportion of phishing immune individuals can help those that are not and aid the resilience of the organization as a whole. Preliminary results show how to measure users' mental processes as a way forward to improve the phishing reporting process altogether. We promote this idea using a real world example and provide directions on how to make human reports actionable.

Acknowledgment This work is supported by the ITEA3 programme through the DEFRAUDIfy project funded by Rijksdienst voor Ondernemend Nederland (grant no. ITEA191010).

References

[1] Verizon, "Data Breach Investigations Report," Tech. Rep., 2019.

[2] L. Allodi, T. Chotza, E. Panina, and N. Zannone, "On the Need for New Antphishing Measures Against Spear Phishing Attacks," *IEEE Security & Privacy*, vol. 18, no. 2, pp. 23–34, 2020.

[3] S. L. Blond, A. Uritesc, C. Gilbert, Z. L. Chua, P. Saxena, and E. Kirda, "A Look at Targeted Attacks Through the Lense of an NGO," in *USENIX Security Symposium*, 2014, pp. 543–558.

[4] G. Ho, A. Sharma, M. Javed, V. Paxson, and D. Wagner, "Detecting Credential Spearphishing in Enterprise Settings," in *USENIX Security Symposium*, 2017, pp. 469–485.

[5] A. Cidon, L. Gavish, I. Bleier, N. Korshun, M. Schweighauser, and A. Tsitkin, "High Precision Detection of Business Email Compromise," in *USENIX Security Symposium*, 2019, pp. 1291–1307.

[6] A. J. Burns, M. E. Johnson, and D. D. Caputo, "Spear phishing in a barrel: Insights from a targeted phishing campaign," *Journal of Organizational Computing and Electronic Commerce*, vol. 29, no. 1, pp. 24–39, 2019.

[7] F. B. Kokulu, A. Soneji, T. Bao, Y. Shoshitaishvili, Z. Zhao, A. Doupé, and G.-J. Ahn, "Matched and Mismatched SOCs: A Qualitative Study on Security Operations Center Issues," in *SIGSAC Conference on Computer and Communications Security*. ACM, 2019, pp. 1955–1970.

[8] T. Sommestad and H. Karlzén, "A meta-analysis of field experiments on phishing susceptibility," in *Symposium on Electronic Crime Research*. IEEE, 2019.

[9] A. Aleroud and L. Zhou, "Phishing environments, techniques, and countermeasures: A survey," *Computers & Security*, vol. 68, pp. 160–196, 2017.

[10] A. M. Ferreira and P. M. V. Marques, "Phishing Through Time: A Ten Year Story based on Abstracts." in *ICISSP*, 2018, pp. 225–232.

[11] A. J. Ferguson, "Fostering E-Mail Security Awareness: The West Point Carronade," *Educause Quarterly*, vol. 28, no. 1, pp. 54–57, 2005.

[12] P. Kumaraguru, J. Cranshaw, A. Acquisti, L. Cranor, J. Hong, M. A. Blair, and T. Pham, "School of phish: a real-word evaluation of anti-phishing training," in *Symposium on Usable Privacy and Security*. ACM, 2009.

[13] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," *Commun. ACM*, vol. 50, no. 10, pp. 94–100, 2007.

[14] D. D. Caputo, S. L. Pfleeger, J. D. Freeman, and M. E. Johnson, "Going Spear Phishing: Exploring Embedded Training and Awareness," *IEEE Security Privacy*, vol. 12, no. 1, pp. 28–38, 2014.

[15] P. Kumaraguru, S. Sheng, A. Acquisti, L. F. Cranor, and J. Hong, "Lessons from a real world evaluation of anti-phishing training," in *2008 eCrime Researchers Summit*, 2008, pp. 1–12.

[16] R. Wash and M. M. Cooper, "Who Provides Phishing Training? Facts, Stories, and People Like Me," in *Conference on Human Factors in Computing Systems*, 2018.

[17] W. Arthur Jr, W. Bennett Jr, P. L. Stanush, and T. L. McNelly, "Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis," *Human Performance*, vol. 11, no. 1, pp. 57–101, 1998.

[18] J.-W. Bullee, "Experimental social engineering: investigation and prevention," PhD Thesis, CTIT, 2017.

[19] M. Khonji, Y. Iraqi, and A. Jones, "Mitigation of spear phishing attacks: A Content-based Authorship Identification framework," in *International Conference for Internet Technology and Secured Transactions*, 2011, pp. 416–421.

[20] G. Stringhini and O. Thonnard, "That Ain't You: Blocking Spearphishing Through Behavioral Modelling," in *Detection of Intrusions and Malware, and Vulnerability Assessment*, 2015, pp. 78–97.

[21] S. Duman, K. Kalkan-Cakmakci, M. Egele, W. Robertson, and E. Kirda, "EmailProfiler: Spearphishing Filtering with Header and Stylometric Features of Emails," in *Annual Computer Software and Applications Conference*, vol. 1, 2016, pp. 408–416.

[22] M. Zhao, B. An, and C. Kiekintveld, "Optimizing personalized email filtering thresholds to mitigate sequential spear phishing attacks," in *Conference on Artificial Intelligence*. AAAI Press, 2016, p. 658–664.

[23] A. v. d. Heijden and L. Allodi, "Cognitive Triaging of Phishing Attacks," in *USENIX Security Symposium*, 2019, pp. 1309–1326.

[24] G. Ho, A. Cidon, L. Gavish, M. Schweighauser, V. Paxson, S. Savage, G. M. Voelker, and D. Wagner, "Detecting and Characterizing Lateral Phishing at Scale," in *USENIX Security Symposium*, 2019, pp. 1273–1290.

[25] T. Moore and R. Clayton, "Evaluating the Wisdom of Crowds in Assessing Phishing Websites," in *Financial Cryptography and Data Security*, ser. LNCS. Springer, 2008, pp. 16–30.

[26] E. Fink, M. Sharifi, and J. Carbonell, "Application of Machine Learning and Crowdsourcing to Detection of Cybersecurity Threats," Tech. Rep., 2011.

[27] G. Liu, G. Xiang, B. A. Pendleton, J. I. Hong, and W. Liu, "Smartening the crowds: computational techniques for improving human verification to fight phishing scams," in *Symposium on Usable Privacy and Security*. ACM, 2011.

[28] J. Hong, "The State of Phishing Attacks," *Commun. ACM*, vol. 55, no. 1, pp. 74–81, 2012.

[29] P. Burda, T. Chotza, L. Allodi, and N. Zannone, "Testing the effectiveness of tailored phishing techniques in industry and academia: a field experiment," in *International Conference on Availability, Reliability and Security*. ACM, 2020.

[30] N. A. Jones, H. Ross, T. Lynam, P. Perez, and A. Leitch, "Mental Models: An Interdisciplinary Synthesis of Theory and Methods," *Ecology and Society*, vol. 16, no. 1, 2011.

[31] C. Bird, "Interviews," in *Perspectives on Data Science for Software Engineering*, pp. 125–131.