



REVIEW

Open Access



Don't click: towards an effective anti-phishing training. A comparative literature review

Daniel Jampen^{*}, Gürkan Gür^{}, Thomas Sutter^{} and Bernhard Tellenbach^{}

^{*}Correspondence:
jamp@zhaw.ch
Institute of Applied
Information Technology
(InfT), Zurich University
of Applied Sciences (ZHAW),
Winterthur 8401, Switzerland

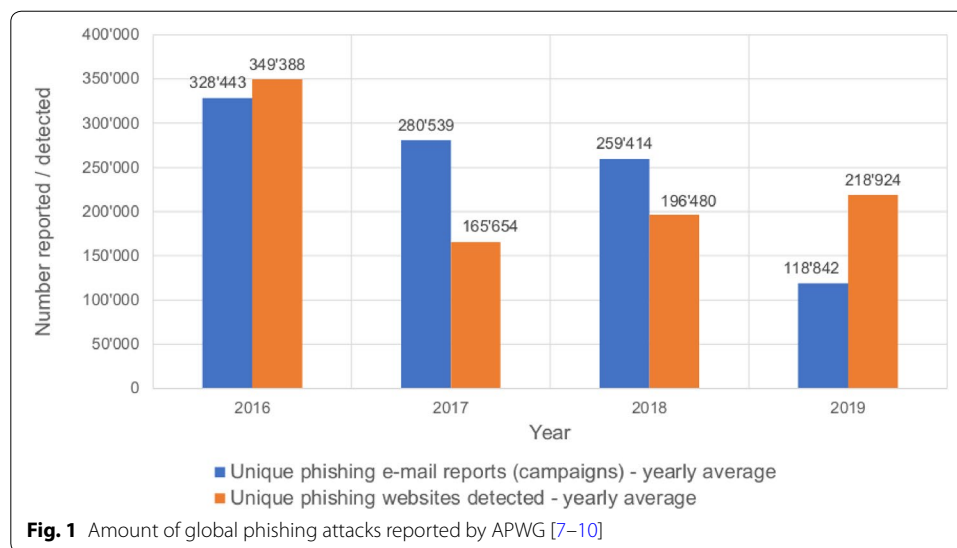
Abstract

Email is of critical importance as a communication channel for both business and personal matters. Unfortunately, it is also often exploited for phishing attacks. To defend against such threats, many organizations have begun to provide anti-phishing training programs to their employees. A central question in the development of such programs is how they can be designed sustainably and effectively to minimize the vulnerability of employees to phishing attacks. In this paper, we survey and categorize works that consider different elements of such programs via a clearly laid-out methodology, and identify key findings in the technical literature. Overall, we find that researchers agree on the answers to many relevant questions regarding the utility and effectiveness of anti-phishing training. However, we identified influencing factors, such as the impact of age on the success of anti-phishing training programs, for which mixed findings are available. Finally, based on our comprehensive analysis, we describe how a well-founded anti-phishing training program should be designed and parameterized with a set of proposed research directions.

Keywords: Phishing, Phishing countermeasures, Anti-phishing training, Security awareness, Security training tools, Machine learning

Introduction

The security threat posed by email-based phishing campaigns targeted at employees is a well-known problem experienced by many organizations. Attacks are reported each year, and a reduction in the number of such attacks is unlikely to occur in the near future (see Fig. 1). A common type of phishing attack involves an attacker attempting to trick victims into clicking on links sent via email. Such links redirect victims to websites that are carefully designed to mimic those of legitimate organizations with the goal of convincing users to provide their personal information and credentials. Attackers then use the *phished* data to execute their schemes further. Phishing attacks may be used to obtain access to an organization's internal servers and steal company secrets or to steal victims' personal information, such as credit card details [1]. In this publication, we focus on email-based phishing attacks, as this is currently the most commonly used channel and poses a significant threat to both individuals and companies globally



[2]. Therefore, in this paper, the term phishing always refers to email-based phishing. Phishing is a lucrative criminal activity that is seldom prosecuted. Moreover, take-down measures are often ineffective, as the landing pages used in phishing attacks transmit the stolen data before they can be shut down [3]. As depicted in Fig. 1, the amount of global phishing attacks is still huge despite more efforts in combatting them. Failing to address or ignoring the threat posed by phishing can result in detrimental consequences for any company. The 2015 Sony Inc. hack is an example of a successful phishing campaign and demonstrates the extent of the damage that such an attack can cause [4]. In this case, according to the *New York Times*, the damage was in the order of hundreds of millions of US dollars [5].

To increase the perceived legitimacy of phishing emails, attackers often adjust their campaigns according to current events. For example, shortly after the publication of the results of the 2016 United States (US) election, Russian hackers began sending emails with malicious zip files attached from spoofed Harvard University email addresses allegedly explaining “Why American Elections are Flawed” [6]. Thus, phishing attacks can be very organized and sophisticated, with the potential to cause extensive damage to the targeted party and maximize the gains for the attackers. The damage caused by phishing attacks can only be estimated, as not all incidents are reported, and the overall damage caused can be challenging to quantify [2]. Nevertheless, Hong et al. [11] reported that the direct loss caused by phishing in the U.S. varies from 61 million to 3 billion USD per year. However, these figures do not reflect the whole picture, as substantial indirect costs are also incurred of post-attack disruption to the ordinary course of business. In addition, phishing attacks are often used as a starting point for other detrimental cyberattacks [4]. As stated in the *2019 Ninth Annual Cost of Cybercrime Study* published by Accenture Security, attackers often begin by targeting the human layer, which is the weakest link in corporate electronic security [2]. In 2013, the *Wall Street Journal* published an article estimating the annual cost of cybercrimes in the U.S. at 100 billion USD [12]. Similarly, based on the rapid global digitalization of consumers’ lives and enterprise records, Juniper Research estimated the costs resulting from data breaches in 2015,

reaching 500 billion USD globally [13]. The recent *2019 Official Annual Cybercrime Report* from Cybersecurity Ventures, discusses costs of up to 3 trillion USD globally for 2015 and estimates that this figure will double by 2021 [14]. Despite the variety in terms of the figures estimated, the resulting picture is clear: As a security threat, phishing has to be taken seriously, as it can cause both direct and indirect costs and can open the door to other, even costlier, attacks [15].

Although there exist various technical solutions intended to prevent phishing emails from reaching their targets, such systems are not perfect and cannot filter out all malicious emails [16]. Attackers (i.e., phishers), have invariably found means of circumventing newly implemented protection mechanisms in the long run [17]. In that regard, techniques based on Machine Learning (ML) have yielded promising results compared to other solutions, as, in some cases, they have almost completely defeated zero-hour phishing attacks and have demonstrated very high true-positive detection rates [18]. Nevertheless, ML does not represent a “silver bullet” against phishing as there are practical challenges that remain to be overcome, such as how these systems should be trained or the threat of adversarial use of ML. As ML is still not an entirely bullet-proof technique, the phishing problem continues to pose a threat [19, 20].

Therefore, an essential part of any institution’s anti-phishing strategy is to take a proactive stance by educating its users so that they can identify phishing emails themselves and act accordingly. There are various suggestions concerning how this goal can be accomplished, including offering dedicated courses or simulating encounters with phishing emails, with such scenarios often developed by an institution’s own security staff. As research results show, however, it is unclear how successful any of these methods are (see “[Impact of anti-phishing training](#)” section for a detailed discussion). Thus, considerable academic attention has recently been paid to how anti-phishing education can be improved and how the utility of this proactive approach can be maximized. However, sources in the literature are occasionally not consistent among themselves concerning specific factors and their impact. This phenomenon results in a situation in which it is challenging for practitioners to create efficient anti-phishing training programs based on academic findings. Additionally, to make the identification of research gaps easier, researchers need to make additional efforts to grasp an overview of the current state of the art. Given the lack of consensus in the literature, it remains challenging to create an effective anti-phishing training program.

Research contributions

Designing an effective anti-phishing strategy involves considering multiple factors, such as *how*, *when* and *at what frequency* users should undergo training. In this work, the term *training* is used to refer to a process (e.g., a course), intended to improve a person’s awareness and knowledge of phishing, which in turn has a potential impact on his or her ability to detect and respond to phishing attempts. Such training can involve different instruments or media, such as computer-based simulations, videos, and leaflets or other printed materials.

To address the challenges associated with training employees to avoid such attacks, we identify relevant factors that should be considered in a company’s anti-phishing training program, then provide a comprehensive survey of relevant research results and, based

on these findings, present a proposal for an ideal anti-phishing training program. The research questions we address within this survey are:

- What are the relevant factors of an effective anti-phishing training?
- Are there any controversial reports of anti-phishing training effects in the academic work of the last 17 years?
- What are the implications of current research findings for designing effective anti-phishing training programs?

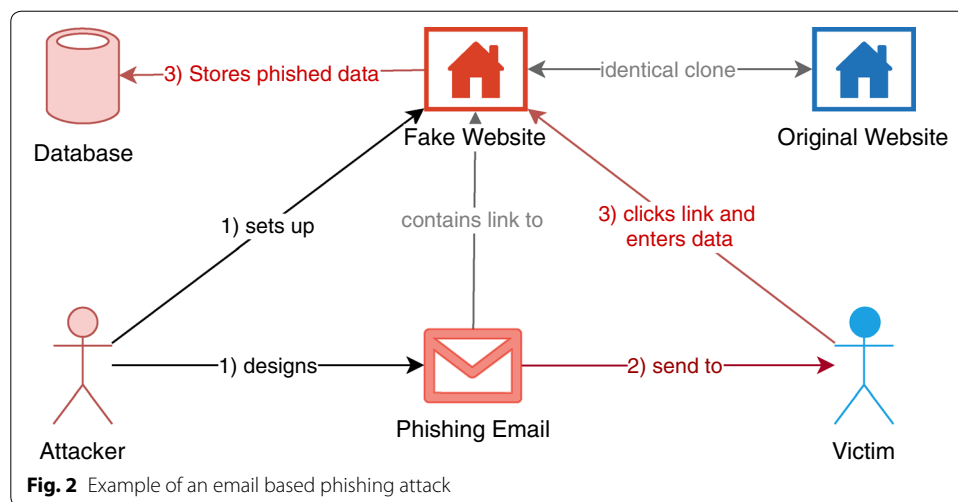
This effort is crucial, as insights into anti-phishing training and into how an effective training program can be developed are instrumental in improving defense against phishing attacks. Moreover, a training program serves to reduce potential damage and increase the overall security of organizations. Current research indicates that factors such as the selected training method, how feedback should be provided to users, how training materials should be designed and how retraining intervals should be organized are relevant and thus have direct impact on the success of an anti-phishing program [21–26]. Considering these findings, this paper makes the following contributions:

- It identifies relevant academic works on anti-phishing training (“[Methodology](#)” section);
- It defines multiple categories, each covering one or several of the identified core areas by examining and categorizing the surveyed works (“[Categories](#)” section);
- It concisely presents the most important findings of each study and their implications for an envisaged training program (“[Literature analysis](#)” section);
- It proposes an effective anti-phishing training program based on the performed analysis (“[Discussion](#)” section).

The next section provides essential background information on phishing. We briefly discuss what phishing is and what can be done to address it. In “[Methodology](#)” section, we describe the methodology applied in the identification and categorization of phishing studies. We then present a comparative literature analysis, which includes a detailed discussion of findings from a wide range of research works in “[Literature analysis](#)” section. The discussion in “[Discussion](#)” section further elaborates on those findings to indicate how they can be used to improve the design and execution of an effective anti-phishing training program. “[Discussion](#)” section also describes how anti-phishing training tools intended to support the features required for such programs should be developed. Then, “[Conclusion](#)” section presents our conclusions and key findings, followed by future research directions in “[Future research directions](#)” section.

Phishing

The term *phishing* refers to attempts by attackers to trick victims into performing a specific action. The objective of such an action could be manifold: it may aim to make the users click on an email attachment, download and execute a file from the Internet. It may also trick them to execute an action on an online platform or to unknowingly provide confidential information such as login or bank details [1, 27, 28]. Often, attackers

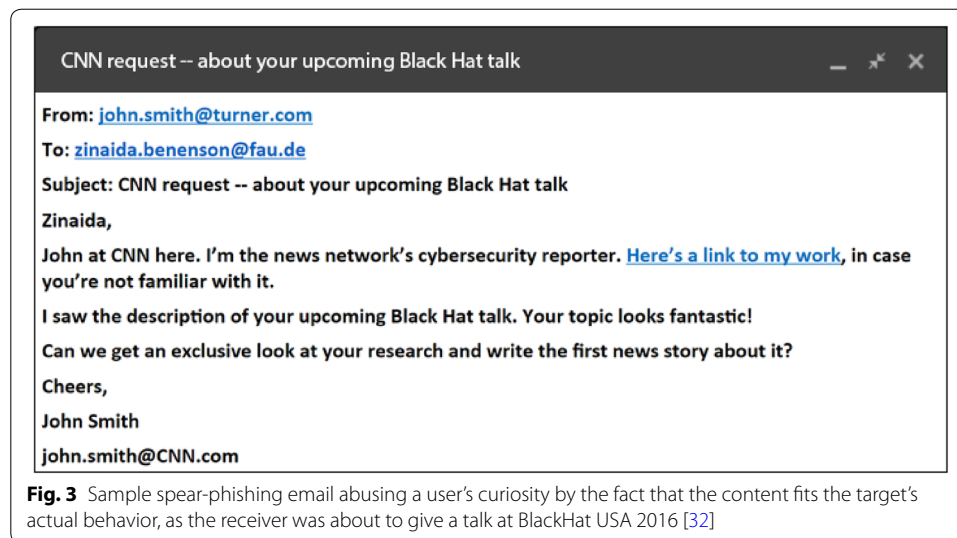


first attempt to gain the trust of their victims and then abuse that trust to lure users into accessing fraudulent or hacked websites. These websites are carefully designed not to make users suspicious and, for example, contain forms that relay entered data directly to the attacker, provide malicious files to download, or contain exploits that infect the victims' devices with malware [11]. Such malware can then be used for a wide range of attacks, such as infecting the target with ransomware or engaging in industrial espionage.

Figure 2 presents an example of an email-based phishing attack [29]. First, attackers identify an existing website containing a form requesting the data they wish to obtain. They then set up a phishing website by cloning the existing one, design an email containing a link to the phishing website (step 1), and send the email(s) to the phishing victim(s) (step 2). In the event of a successful attack, the victim thinks the email is legitimate; he or she then clicks the link and provides an attack with the desired data (step 3).

One key element of an institution's anti-phishing campaign is the education of its employees. Currently, several techniques are used to achieve this goal. The most prominent of these are providing informative material concerning phishing, offering dedicated computer-based or "offline" anti-phishing training courses, and developing a phishing simulation that provides anti-phishing training materials if a link is clicked [21, 22, 25, 30]. All of these strategies have different pros and cons, which should be considered from a cost-benefit perspective. In particular, the resources required (e.g., money) are often a decisive limiting factor. For instance, according to research based on responses provided by 500 companies with between 1000 and 5000 employees in 2017, the cost of user security education that includes anti-phishing training has reached approximately 290 K USD per year for large enterprises [31]. Therefore, we believe that determining the most effective training program is critical in overcoming such concerns and improving cybersecurity. However, please note that addressing the cost-effectiveness of training programs is beyond the scope of this survey.

Besides generic *phishing*, in which many potential victims are targeted, *spear phishing* describes a specially tailored phishing attack against one victim or a small group thereof [33]. Attackers research their targets and abuse the acquired information to design



phishing emails for each victim. As, in such cases, the contents of a phishing email will then reflect the target's current activities, the likelihood of the attack succeeding will probably increase. An example of a spear phishing email exploiting the current activities of its target is presented in Fig. 3.

Phishing countermeasures

Phishing countermeasures can be applied at several stages during an attack. Considering the general attack model in Fig. 2, a technical **filtering solution** could be deployed in Step 2 (e.g., [20]). Such solutions process and analyze all incoming email messages and, based on rules or ML, classify them as either *phishing* or *legitimate*. ML filtering techniques have become state of the art and the classification of phishing website (e.g., [34–38]) can be used for blacklists. Such approaches can prevent a phishing message from reaching the target user, but attackers can use ML techniques as well (e.g., [39]) for bypassing such AI detection systems. Furthermore, ML based countermeasures can be further adapted and optimized for different operational environments to improve performance and combat implementation challenges. A recent example is [40] where the authors have implemented an anti-phishing virtual network function at the edge of the network with embedded robust machine learning techniques for phishing detection.

Alternatively, **education of users** is a proactive method. In other words, users themselves could be educated to identify phishing scams (e.g., [41]). By creating awareness of phishing attacks and training users to be able to identify them, this method can prevent employees from falling victim to phishing scams and therefore prevent possible information leaks. Additionally, **web filtering software or a specific firewall** could be used to analyze all of the websites visited by an employee (Step 3 in Fig. 2) and attempt to prevent access to sites with malicious intent (e.g., [42]). This would again prevent users from inadvertently leaking information. A more active approach is the **take-down of phishing websites** by third parties such as the law enforcement agencies (LEAs) or the hosting services to prevent potential victims from accessing such websites (e.g., [3]). According to Hong et al. [11], the duration of the entire

Table 1 Phishing countermeasures

Counter-measure	Example	Approach
Filtering	[20]	Analyze all incoming email messages and filter them based on their class: phishing or legitimate
Education of users	[41]	Train employees to protect them from falling victim to phishing attacks and prevent possible information leaks
Web filtering	[42]	Analyze browsed websites and identify possible phishing sites in order to warn the user or completely block access to the suspected sites
Website take-downs	[44]	Take down the phishing site to prevent potential victims from accessing such websites (usually by external parties such as LEAs or hosting service providers)

take-down process averages approximately 62 h. An important requirement for an efficient mitigation effort is multi-agency participation (e.g. Internet users, brand enterprises, browser manufacturers and authorities) with uniform data sharing format and unobstructed sharing channels for common phishing reporting. One way to achieve this is with multi-party phishing data sharing platform based on blockchains [43]. In summary, the methods listed in Table 1 are available to counteract phishing.

When attempting to address phishing attacks, security should always be implemented using multiple layers of defense (defense-in-depth), as each layer has its strengths and weaknesses [45]. In that regard, each layer should be considered breakable, as no bullet-proof solution against phishing currently exists. Therefore, a combination of the layers, as mentioned earlier, would be an approach to the problem. An essential aspect of such a defense strategy would be to educate employees and strengthen their ability to identify phishing attacks. This requirement raises the critical question of how anti-phishing training programs and tools should be designed and implemented, which constitutes the primary rationale for the contributions made by this paper.

Methodology

In this section, we describe how the literature for this survey was selected. We explain the methods used for searching, filtering, and selecting the literature. Moreover, we introduce a categorization system for academic anti-phishing training papers and use this system in “Literature analysis” section to categorize the selected papers.

During this study, we carried out two iterations of our literature selection process. We conducted the first iteration in November 2018, and it includes articles from 2003 to 2018. It contains the main corpus of our research. In April 2020, we executed the second iteration of our selection with the scope of articles from 2019 and 2020 during the peer review process of our paper. We chose articles between 2019 and 2020 for the completeness of our survey and because we wanted to include the latest state-of-the-art articles when it is published.

The literature related to anti-phishing training is extensive. It covers areas ranging from technical approaches for exploiting a weakness of a given email client with phishing purposes to user education in general. In this work, we analyze a comprehensive set of publications related to factors relevant to anti-phishing training and the

Table 2 Methodology to search, filter, and select the literature

Selection step	Articles 2003–2018	Articles 2019–2020
Step 1: Search using Google Scholar (GS) using the keyword “phishing” and a filter for the time frame. Only the first 1000 search results were considered	37,300	13,000
Step 2: Remove patents, duplicates, and publications without publisher or source	726	607
Step 3: Determine whether the content is usable by considering publication titles and, if non-conclusive, abstracts, and conclusions	81	67
Step 4: Apply paper quality control filters	77	47
Step 5: Manually check the remaining publications for valuable content	67 ^A	37 ^B
Total number of publications included in the survey:	^A + ^B	→ 104

success thereof. There are several types of training, such as the use of using videos, web-based courses, informational material such as leaflets/flyers, or simulated phishing attacks. No specific type of training is favored in this publication.

To this end, “[Identifying relevant sources](#)” section provides information concerning how the works included in this survey were chosen (i.e., what criteria they had to meet). In “[Categories](#)” section, we accordingly construct categories based on the identified literature, which we later utilize to classify the surveyed papers.

Identifying relevant sources

To render it as reproducible and as clear as possible, we have divided the identification process into multiple steps. This section explains each step in detail, while Table 2 shows how those steps were used to narrow the relevant literature down to the works included in this survey.

Prior to the first step, in which the search keywords were defined and the initial set of publications was obtained, potential electronic sources were evaluated. Querying the search engines of IEEE¹, ACM², ScienceDirect³, Wiley⁴ and GS⁵ with the keyword “anti-phishing training” (including the quotation marks) returned the following number of publications: *GS*: 406, *IEEEExplore*: 2, *ACM Digital Library*: 3, *ScienceDirect*: 11, *Wiley Online Library*: 5. Cross-checking the publications returned by each search engine indicated that the search engine that returned the most results, GS, already included the publications found by the other engines. As stated in the GS *About* page [46], this engine provides a service that allows users to search the databases of many publishers from one location as reflected in the results of our initial search engine test. Therefore, GS was selected as the literature search engine for this survey.

In *Step 1* of the first iteration, which is the start of the literature identification process, GS was used to obtain a set of publications for potential inclusion in this survey.

¹ IEEEExplore: <https://ieeexplore.ieee.org>.

² ACM Digital Library: <https://dl.acm.org>.

³ ScienceDirect: <https://www.sciencedirect.com>.

⁴ Wiley Online Library: <https://onlinelibrary.wiley.com>.

⁵ GS: <https://scholar.google.com>.

Table 3 Overview of the number of participants in surveyed publications, grouped into the ranges very large, large, medium, small and not applicable

Name	Number of participants	Amount	Publications
Very large	$\geq 10,000$	6	[48–53]
Large	1000–9999	18	[30, 54–70]
Medium	100–999	32	[21, 22, 25, 26, 71–98]
Small	20–99	23	[41, 99–120]
Not applicable	–	25	[11, 16–18, 20, 23, 24, 121–138]

Not applicable means that no experiments were performed. For example, such a paper could be a survey

In the initial search engine test, the keywords “anti-phishing training” has been used. As this query returned only 406 publications, we expanded the search by using a more general keyword: “phishing”. We ensure that the publications remain relevant with a 15-year date filter (2003 to 2018) that we added to the query. This search returned 37,300 results, of which GS returned the first 1000 publications, sorted by relevance. GS ranks the list of results by weighting the full text of each document, where it was published, by whom it was written, and how often as well as how recently it has been cited in other literature [46]. As an additional check, the other search engines queried in the initial search test were also queried using the new keyword, whereupon *IEEEexplore* returned 1040 publications, *ACM Digital Library* 1148, *ScienceDirect* 2678 and *Wiley Online Library* 1241. As GS does not provide a way to export search results, the tool *Publish or Perish* was used to perform this task [47].

Step 2 involved the application of a generic filter to the data set obtained in *Step 1* to focus on scholarly publications with complete identifier data. In that regard, we removed publications for which GS could not identify a publisher or source. Moreover, we applied a filter that removed patents.

In *Step 3*, we used the publication title as the primary indicator of whether it is relevant to the focus of this paper. First, we conducted a keyword search with the following keywords: Awareness, training, phishing, susceptibility, and behavior. In case one of the keywords was found within the title, we directly selected it for *Step 4*. If we did not find any of the keywords, we examined the abstract and the conclusion of the paper. If the abstract or the conclusion of the paper had a focus of anti-phishing training, we selected it for *Step 4*. The remaining steps constituted a progressive identification process (elimination according to various attributes such as being peer-reviewed and the level of relevance) for the list of publications.

In *Step 4*, we applied an attribute-based approach to filtering to the list from *Step 3* for quality control in terms of peer reviewing and experiment design. We applied the following quality criteria in this step:

1. Peer-reviewed: To ensure the quality and reliability of our survey’s conclusions, only peer-reviewed papers are considered
2. Target study group size: Works presenting conclusions based on a low participant number n in their studies ($n < 20$) are not included (see Table 3 for further details)

Table 4 Overview of all the publications analyzed in this survey

Category	Amount	Publications
Training impact	35	[16, 21, 22, 24, 30, 41, 50, 51, 53, 55, 58, 60, 61, 67, 69, 71, 72, 75, 76, 80, 82, 91, 92, 99, 101, 102, 104, 107, 109, 110, 114, 116, 122, 125, 126]
Target group impact	53	[21, 25, 30, 48, 49, 52, 54–56, 59, 65, 66, 70, 72–75, 77–86, 88–90, 93, 94, 97–101, 103–107, 112, 113, 115, 116, 118, 125, 127, 131, 136, 137, 138]
Email content and structure	24	[17, 18, 50, 55, 62, 63, 68, 79, 84, 87, 91, 95, 96, 111, 119, 120, 124, 127–130, 133, 134, 136]
Feedback	23	[21–26, 30, 41, 50, 55, 57, 62, 64, 75, 82, 84, 101–103, 108, 121, 132, 135]
Knowledge retention	12	[21, 23, 24, 55, 57, 61, 69, 75, 82, 109, 110, 128]

3. Control group: All publications that involve actual participant training have to make use of a *control group* to verify their findings against participants who do not undergo any training
4. Language: Only publications in English are considered

The final step, *Step 5*, requires the most effort, as it is in this stage that the list of publications identified in *Step 4* are thoroughly evaluated and selected due to their merit. In this step, each remaining paper was examined, read, and reviewed to determine whether it offers relevant contributions to the focus of the present study.

For the second iteration, we applied the same steps, but for the time between 2019 to 2020. Finally, the merger of the two result sets provided us the surveyed paper base.

Limitations of survey methodology Identifying studies for potential inclusion in a literature survey is a process wherein limits and boundaries have to be set carefully. Depending on the adopted data collection guidelines, a comprehensive overview of the existing literature should be provided in survey work, although such an overview is usually not exhaustive in terms of coverage. For our work, GS was used as the search engine of choice. Since the engine already returned many relevant publications (e.g., 37,300 results for 2003–2018 period), no additional searches (e.g., backward/forward citation searches) were performed. Such complementary methods may have led to the identification of additional articles; thus, some relevant articles may have been omitted due to the methodology in the present study, which was based on keyword searches. However, considering the vast body of articles evaluated and the fact that the rating system of GS considers both relevance and impact, we believe that this work presents a comprehensive study and contributes to research on anti-phishing training.

Categories

We divide the selected publications into multiple categories, each of which covers one or multiple core area(s), which were identified by examining those works. A paper can appear in multiple categories should its results cover more than one area/factor. For a better overview of the approach to categorization described in the following sections, all papers, along with their corresponding categories, are presented in Table 4. We compare the results of a reviewed paper with those of other studies in the same category. Therefore, the objective of the comparative analysis conducted in this paper is to *identify contradictory findings and evaluate consistency* with other findings and conclusions.

Research results concerning phishing mostly address either *attack success rates* or *training effects*. For example, a work stating that emails containing links in a specific format are more successful would fall into the former group while another describing an educational game and the effects thereof on the participants would belong to the latter. In our categorization, the second and third categories in Table 4 include papers concerning attack success rate, while the others address training effects. Both of these groups are crucial since they render the inherent factors on anti-phishing training design and effectiveness by jointly illustrating aspects of susceptibility, attack success, and training efficiency.

Our classification does not cover all possible relevant features of an anti-phishing campaign. In the context of this survey, we identify the dominant ones and restrict the analysis to them in the interests of concision and clarity. We list all categories and provide brief descriptions of the data that are covered by them:

- **Training impact** contains data concerning the training effects after exposing users to anti-phishing training. The key questions are related to the benefits of training: Is educating users a viable approach? Does it help at all, or should the focus of IT personnel be on other phishing defense strategies?
- **Target group impact** contains findings regarding which users exhibit a better or worse ability to identify phishing threats and whether this ability changes as a result of anti-phishing training. This data can be used to find weak links within an organization and direct increased training efforts towards strengthening them.
- **Email content and structure** includes data related to the design and structure of phishing emails. The essential questions concern the visual appearance of such emails and how it is designed, the content and how (URLs) can make it easier or more difficult for users to determine the legitimacy of a phishing email.
- **Feedback** contains information concerning the design of the web pages for phishing attacks and the phishing training material. Moreover, it contains findings of possible ways to present educational material and how effective these are.
- **Knowledge retention** includes findings of how long the knowledge gained through educational measures is retained, how effective it is, and in which intervals users should be retrained.

Literature analysis

In this section, we present the analysis of surveyed works structured into the categories defined in our methodology. In each category, we employ a consistent approach to structuring our review of the literature: A discussion of related works follows a short introductory section; after that, a concise analysis of the findings of each study is performed to identify any common traits and to draw conclusions. This presentation pattern is intended to make the literature review and analysis more accessible for the reader.

Impact of anti-phishing training

Introduction

A fundamental question concerning the design and structure of training programs is whether or not they have a measurable impact on employee behavior with regard to

phishing emails. In the following survey segment, we first discuss those studies that report a positive effect of anti-phishing training measures. We then conclude the survey by reviewing literary sources that report mixed results. As an important note, the list of papers obtained through the selection process does not contain any papers that report negative results only.

Survey

There is a large body of publications that confirm a decreased likelihood that users will fall victim to phishing messages after educating them with general anti-phishing material or via embedded training. The latter is a training method that is seamlessly integrated into a user's typical workflow. Embedded training usually works along the following lines: A training system sends artificial phishing emails to a set of trainees. The trainees have to identify and report those emails when they process their emails during a typical workday. In contrast to controlled training environments, embedded training occurs under realistic settings in which trainees might be affected by factors such as distractions, stress, and a lack of focus. Should a trainee click on the link in the phishing email, he or she will fail the training and receive some kind of education (see "[Feedback](#)" section).

In [104], Neupane et al. conducted a multi-modal neurophysiological study regarding phishing detection and malware warnings. In regards to the participants' trainability, the authors found that their users were paying attention to the information provided and made active efforts while performing the assigned tasks. Based on these results, the authors conclude that the participants in their study did not ignore provided training materials and that training is indeed a valuable approach to address phishing. In the same vein, Halevi et al. confirmed that awareness, which often increased due to training, helped their participants not to fall victim to phishing as the subjects were more concerned about protecting themselves [107]. In [101], Greene et al. examined long-term, operationally-situated data that was captured during embedded phishing awareness training exercises held throughout four and a half years at a U.S. government institution. Apart from an improved phishing detection rate, the authors also observed new competition due to the gamification of the phishing awareness training exercises over the years. Participants would attempt to beat their colleagues and be the first to identify the phishing emails, which possibly improved the training results further. Doge et al. [71] report similar success when using embedded training. In an experiment with three groups of approximately 300 participants each, the first group was exposed to embedded training, the second group received a notification after falling victim to a phishing email, and the third was the control group which was not exposed to training. Their results indicate that over a period of 10 days, there was no significant difference in terms of susceptibility among the three groups. However, over a more extended period (63 days in this experiment), training was found to result in significant improvements for the participants' clicking behavior. Of the participants who received training, 24.5% failed the experiment. Of those participants who received feedback alone, 32.08% failed, and, in the group that received neither feedback nor training, 47.5% failed.

A more recent work by Gordon et al. is a retrospective study of employee susceptibility at six US health care institutions. In this multicenter study, phishing simulations (95

campaigns) were run from 2011 to 2018, with 3 million phishing emails sent to employees of those organizations. Overall click rates varied by institution but were notably high: on average, around 400,000 (14%) of simulated emails were clicked on by employees. In their work, repeated phishing campaigns were associated with reduced odds of clicking on subsequent phishing emails. In models adjusted for several potential confounders, including year, the institutional campaign number, institution, and email category, the odds of clicking on a phishing email were 0.511 lower for 6 to 10 campaigns at an institution and 0.335 lower for more than 10 campaigns at an institution. They also found that there were important institutional differences in click rates, as well as differences in click rates between email category and season. Other papers included in this study that report the embedded training method having a positive effect are [21, 22, 24, 41, 82, 99, 109, 110].

Papers that report mixed results but that are partially in favor of a positive effect of anti-phishing training are [50, 55, 72, 76, 80, 126]. In [76], Orunsolu et al. examined the effectiveness of the security tips provided by a Nigerian bank to their customers as a form of education. These messages provide information on how users can identify online scams and which actions users should avoid. The authors' findings showed that most participants were unable to reliably identify a phishing email despite having been exposed to the security tips. After this test, the authors performed a course-based training session, and, in the follow-up test, participants exhibited an increased success rate in identifying phishing threats. In [55], Caputo et al. obtained mixed results in their study regarding the impact of anti-phishing training. They found that the phishing detection rate of members of two groups, whom they referred to as "all clickers" and "non-clickers" did not improve at all, as they always (11%) or never (22%) clicked, regardless of the applied training method. Additionally, the authors grouped the remaining 67% into a group they called "inexplicable." Users in this group seem to click or not click on phishing links randomly. Nevertheless, the authors note that phishing messages that are not detected by technological solutions are often identified as a result of company personnel reporting an email as being suspicious. According to Caputo et al., providing a reporting feature should be considered as a possible additional layer in a company's phishing defense system, mainly, as early reports provide meaningful benefits for members of an organization's incident response team. Karakasiliotis et al. [126] conducted a study to assess end-user awareness of social engineering and phishing. They conclude that a need for increased security awareness is evident but designing a generalized approach to achieving such awareness could be a complicated process due to the technical unfamiliarity of users or behavioral differences among them.

Vishwanath et al. [80] developed a methodology for determining why so many users fall victim to phishing and why this seems to occur on a random basis. They report that a user's susceptibility to phishing depends on multiple factors, only one of which can be trained using the embedded training technique. However, using the method developed by the authors, security officers can identify the weak links within their organizations; in addition, it enables them to determine how much training an employee requires and to set the focus of the training. Siadati et al. [50] found that training participants using persuasive phishing emails significantly improves their average resilience to such emails. In contrast, training involving emails that were not considered to be very persuasive had

little impact on the phishing susceptibility of the investigated users. Finally, Moody et al. [72] report that even with education, users are still overconfident in their ability to detect phishing messages. This overconfidence can, however, be diminished through education.

Summary

As multiple research studies show, an increased ability to correctly handle phishing emails after receiving anti-phishing training is well supported in the scientific community. However, Caputo et al. identified two groups of users who were not affected by the applied training: those who clicked all links and those who never clicked. However, the authors did not address how these groups should be educated.

While most studies have attempted to answer the question of whether training makes trainees less susceptible to phishing attempts, little information is available concerning how such training changes their behavior regarding benign emails. Three notable exceptions are [30, 75, 102]. In [75], Kumaraguru et al. report that embedded anti-phishing training does not affect users' willingness to click on links in benign emails [75]. However, this is in contrast with the findings of Sheng et al., who report that some users stopped clicking on legitimate links in emails when the design of the provided training materials did not take such behavior into account. Unfortunately, the authors did not identify the type of design that could achieve this outcome [30]. The finding of Sheng et al. is confirmed by Yang et al. [102], which confirms that this issue should be given special attention.

Another intriguing issue is raised by a literature survey conducted by Khonjii et al. [122]. They conclude that user education has a positive impact, but they criticize the fact that none of the reviewed studies evaluates whether such improvement is still meaningful when considering different technical phishing-detection solutions. If there are solutions that can filter all but those emails with which users struggle to identify, training would not provide any benefits even after appropriate anti-phishing training.

In summary, these mostly positive results indicate that anti-phishing training indeed has a positive impact. However, training design, especially complementing embedded training with standard training sessions and even individualization of training, might also play an important role.

Target group impact

Introduction

Findings in this category feature works that are related to user-specific properties. For example, they may note that users working in technical jobs are as likely to fall victim to phishing as others. Such insights are critical for identifying groups of users who are more susceptible to phishing. Employees in such groups could accordingly receive additional training or receive different types of training to mitigate possible attacks.

In this part, we focus on the properties presented in Table 5, and we use it as a guide for the discussion of the works considered in this section. More specifically, we first discuss all of the works included in the column titled *Has impact* and then those included in the *No impact* column. Within a column, we start with the papers listed for the first parameter and then continue row by row. However, as most papers present findings concerning more than one parameter and discussing the same paper in multiple places

Table 5 User-specific properties and their impact on susceptibility to phishing attacks

Parameter	Has impact	No impact
Age	[30, 52, 59, 65, 75, 98]	[21, 48, 99]
Gender	[30, 52, 56, 59, 78, 100, 105, 107]	[21, 25, 48, 54, 65, 66, 99]
Technicality	[59, 65, 66, 70, 88, 93, 100, 125]	[81, 82]
Extrovert personality	[125]	–
Known sender address	[72, 131]	–
High use of online activities	[72]	[49]
Email experience	[80, 83–85]	–
Submissiveness	[73, 83]	–
Awareness level	[48, 70, 74, 77, 80, 84, 86, 107]	[66]
Conscientiousness	[90, 107, 125, 131]	–
Attention control/impulsivity	[72, 84, 104, 127]	–
Trust in technological solutions	[74, 101]	–
Risk awareness	[70, 72, 89, 90, 94, 97, 100, 112]	–
Confidence	[74, 86, 90]	–
Willingness	[86, 100, 105, 106, 115]	–
Commitment level	[73]	–
Distrust/fear/anxiety	[106]	[113]

Statements in the remarks column are our interpretation of the documented *impact/no impact* situations

makes little sense, we also discuss the findings related to other parameters on the first mention of a source. As a consequence, when we follow the order of the parameters, only papers that have not yet been introduced will be discussed.

Survey

Papers reporting impact In [75], Kumaraguru et al. report on an experiment in a university setting. They find that participants between the ages of 18 to 25 are consistently more vulnerable than other age groups. Sheng et al. confirm the same finding in [75] regarding this age group. Furthermore, Sheng et al. state that the results of their role-playing online survey instrument-based study involving 1001 participants suggest that women are more susceptible to phishing, probably because they have undergone less technical training [30].

Another study that points in the same direction as Sheng et al. is by Jagatic et al. [56]. The authors tested students and found that a phishing mail was slightly more likely to be successful when the sender was of the opposite gender to the receiver. In [78], Iuga et al. consider relationships between the demographic characteristics of individuals and their ability to correctly identify a phishing attack, as well as the impact of time-related factors. Their results suggest that gender and the number of years of computer usage experience have a statistically significant impact on the phishing detection rate; the same can be observed for the psychological anchoring effect.

Halevi et al. [107] studied the impact of gender, awareness of cyber-risks, and personal traits on spear-phishing susceptibility. They used a combination of a questionnaire and a real-world phishing simulation and found that women are more likely to respond to spear-phishing messages about winning a prize than men and that people who are more

aware of cyber-risks are less susceptible to such attacks. Concerning personality traits, Halevi et al. found that less suspicious/aware online users are more likely to fall victim to phishing and that conscientiousness can be targeted by attackers to gain a higher phishing response rate. They suggest that, based on their findings, a user-targeted approach to phishing defense may be required.

Flores et al. [100] conducted a study with a focus on targeted phishing attacks. Their results contradict the previously discussed findings concerning the impact of gender as they found that women are less susceptible to phishing attacks. Furthermore, the authors report that an individual's trust and risk behavior significantly affected his or her actual behavior during the phishing experiment. Specifically, computer experience at work and willingness to help showed a significant correlation with the participant's phishing susceptibility.

Hong et al. [105] aimed to identify user profiles that can be used to predict when phishing attacks will be successful. They sought to identify attributes that make some individuals more vulnerable to phishing attacks than others. Their results suggest that gender, trust, and personality are among those attributes.

Another user-specific property is the technical background of a person and the degree to which his or her job is technical. Butavicius et al. conducted two experiments: In the first, they did not tell the participants to be aware of phishing emails, whereas, in the second, they did [125]. They found that computer-savvy participants were more vulnerable to phishing attacks; however, this was only found to be the case in the informed experiment. In the non-informed experiment, they performed similarly to the other participants. Also, by comparing their results with those of a prior personality test [139], Butavicius et al. found that participants in the non-informed experiment performed better in terms of detecting phishing emails when they had more extroverted or open personalities. The same was found for less impulsive people in the non-informed experiment. The authors inferred that those participants who probably deliberated over a phishing email appeared to demonstrate better performance in detecting phishing emails [125]. More support for the impact of a person's technical knowledge comes from Flores et al. [70], who investigated the correlation among selected psychological and demographic factors.

Furthermore, to assess the impact of national culture on these correlations, they performed an experiment involving 2099 employees of nine organizations in Sweden, the USA, and India. It was found that general information security awareness, formal information security training, and computer experience showed a positive correlation with phishing resilience. However, the authors also observed that the behavior demonstrated in response to phishing differs among Swedish, US, and Indian employees.

Parsons et al. [88] present another interesting finding concerning the impact of the participant's technical knowledge. They report that whether or not participants are aware that they are participating in a phishing study might have a significant impact on the outcome of such a study. Participants who were informed that they were participating in a phishing study demonstrated significantly better performance in terms of identifying phishing emails and took longer to make decisions. Intriguingly, participants who had formal training in information systems (technicality) performed more poorly overall.

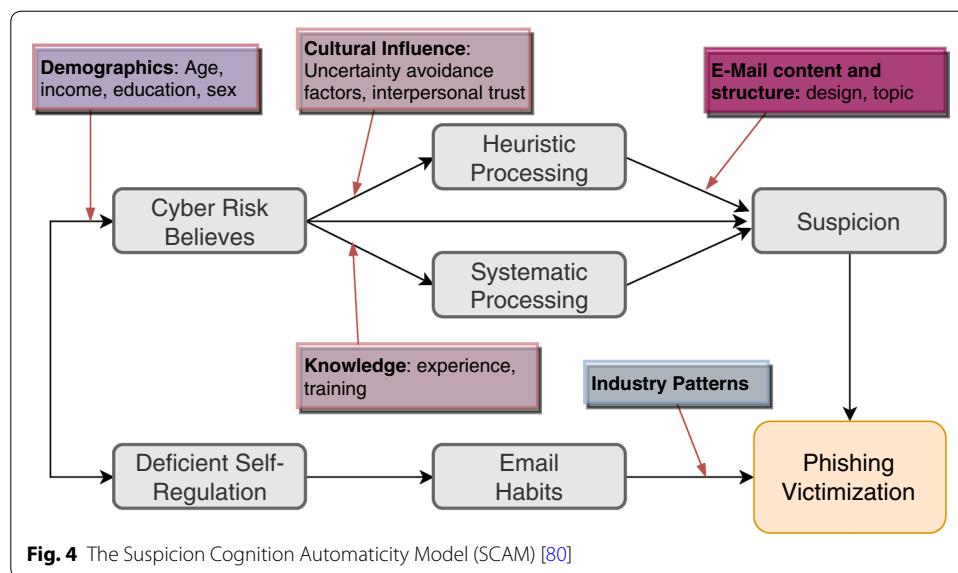
The impact of trust in a sender's email address was the subject of research conducted by Moody et al. [72]. Their results show that users' susceptibility tends to increase when the sender of a possibly fraudulent email is known and reduces if the sender is unknown. The results indicate that users are more likely to click on a link in an email should they believe that the sender is deceitful. This behavior could be caused by the users' desire to discover the true intentions of the sender. Also, the authors state that users who frequently browse the Internet are more likely to click on links in emails than others.

Alseadoon et al. performed a simulated phishing attempt and applied the detection deception model [137] developed by Wright et al. to determine which individuals are more susceptible to phishing. The authors of this study conclude that users who have less email experience and high levels of submissiveness are more likely to fall victim to phishing [83]. Harrison et al. [84] observed that individual factors such as knowledge and experience with email increase resilience to phishing attacks. The focus was on the characteristics of phishing emails, users' knowledge of and experience with phishing, and how these factors interact and influence how users cognitively process phishing emails. It was found that phishing susceptibility can be predicted by a particular combination of a user paying little attention to some aspects of an email and a high degree of elaboration on the part of the phishing message.

However, email experience, especially in the form of personal email habits and processing strategies, might also have a negative impact. Vishwanath et al. [85] compared the causes and consequences of email habits and cognitive processing. The results of their simulated phishing attack indicate that the cumulative effects of heuristic processing and email habits were the main factors affecting the phishing susceptibility, as they were found to cause a fourfold increase in a user falling victim to a phishing attempt and, therefore, nullify any advantage offered by systematic processing [85].

According to the study conducted by Workman [73], people who are more trusting and obedient to authority are more susceptible to social engineering. Furthermore, the author found that people with higher normative, effective, and continuance commitments are more likely to fall victim to phishing attacks. Normative commitment refers to the formation of implied obligations to others. Continuance commitment refers to becoming emotionally invested in a decision, and affective commitment means that people model the behaviors of other groups, role models, or important persons.

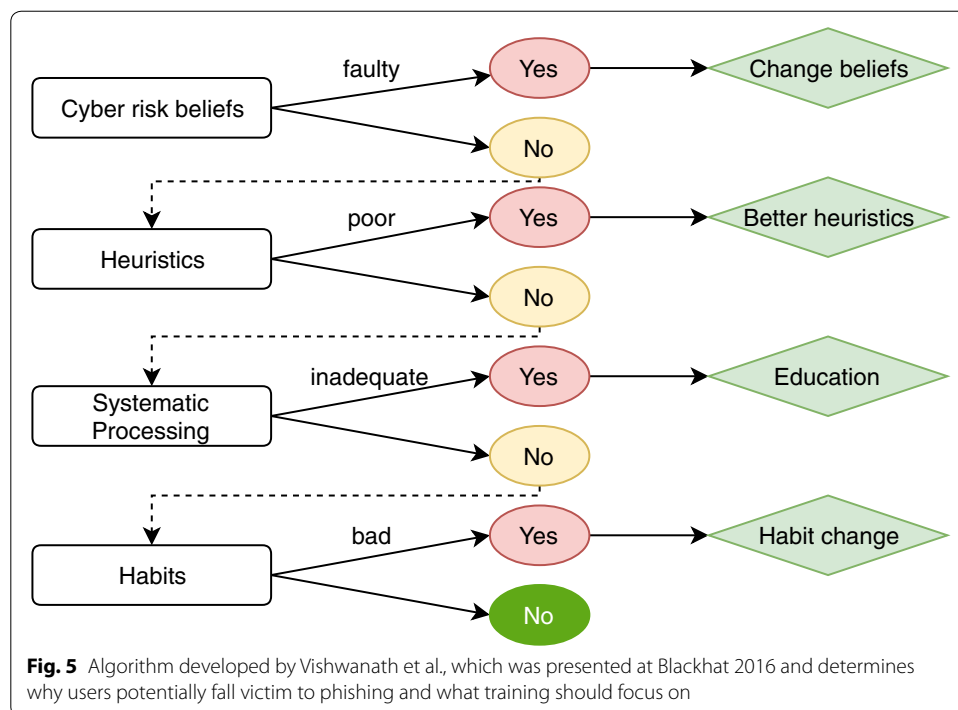
One parameter that is mentioned in many studies is the impact of people's level of awareness. In [77], the authors conducted a phishing exercise in an academic environment as part of an ongoing information security awareness project. They found that educational and awareness activities pertaining to email environments are critical in managing the increased threat of identity theft. Another study pointing in a similar direction is that of [86], in which the authors use signal detection theory to measure vulnerability to phishing attacks, including variation in performance across task conditions. They found that phishing-related decisions are sensitive to individuals' response bias, confidence, detection ability, and perception of consequences (awareness). Specifically, higher sensitivity was found to be positively correlated with confidence, while greater willingness to treat emails as legitimate was negatively correlated with the perceived consequences of participants' actions and positively correlated with confidence.



Arachchilage et al. [112] developed a new game design that educates users about phishing. Their study results showed a significant improvement in participants' phishing avoidance behavior in the second test assessment conducted by the authors. The findings suggest that participants' threat perception, safeguard effectiveness, self-efficacy, perceived severity of a potential threat, and perceived susceptibility elements positively impact threat avoidance behavior, whereas safeguard cost had a negative impact.

Abbasi et al. [74] confirm that awareness is an important factor, but only one of many. Nearly two-thirds of the users in their study fell victim to the phishing mail created by the authors. A cluster analysis of the collected data, which was obtained via questionnaire and phishing simulation, found that, among other factors, over-confidence, a low awareness level, and a high level of trust in technology on the part of the user were detrimental.

In [80], Vishwanath et al., the authors found that a user's awareness level is of similar importance. They observed that research related to human factors and their impact on phishing victimization generally identifies two main sets of factors: The first set is the victim's cognitive processing schema, which is influenced by his or her awareness of the safety of engaging in certain online activities. The second set of factors is the behavior rituals developed by a user based on the work cultures experienced and/or the types of communication devices used. Based on these findings, the authors developed the Suspicion Cognition Automaticity Model (SCAM) model, which is presented in Fig. 4. It describes the likelihood of such victimization of an employee based on the following five parameters: individual beliefs concerning cyber-risk, both heuristic and systematic patterns exhibited while processing an email, deficient self-regulation, and developed email habits. As the SCAM was developed to include all of these parameters, it uses experiential, dispositional, behavioral, and cognitive factors to provide a more comprehensive explanation of the phishing victimization process. Vishwanath presented a proposal concerning how to apply the SCAM for practical use at Blackhat 2016, where he presented how one can calculate the Cyber Risk Index (CRI).



Vishwanath et al. propose a questionnaire with 40 questions, the result of which is used as an input for the algorithm in Fig. 5. First, the algorithm asks the user about his or her cyber-risk beliefs. If these beliefs are faulty, they will have to be changed. If they are reasonable, the user's phishing email identification heuristics are checked. Should a user exhibit poor heuristics, the algorithm will suggest teaching better heuristics. However, should a user apply good heuristics but process emails inadequately, anti-phishing training should be applied. Moreover, even if an employee passes all of these checks, he or she may still fail to identify a phishing email due to bad habits, which would also have to be remedied.

In their multi-modal neuro-physiological study, Neupane et al. [104] found in their multi-modal neuro-physiological study that their participant's personality traits, specifically attention control, directly impacted their phishing detection accuracy. The authors conclude that users may better detect phishing attacks if they could, in addition to undergoing phishing awareness training, be trained to exercise attention control. The authors note, however, that further work is necessary to understand the effect of such interventional training on the user's performance in phishing detection tasks.

In [127], Butavicius et al. point in a similar direction by reporting that the participants in their study who were less impulsive in terms of decision-making were more likely to consider the links in phishing emails as being dangerous. Based on that observation, the authors state that a lower level of cognitive impulsivity could protect against spear phishing. In addition, they found that lower cognitive impulsivity did not adversely influence the participant's judgment of genuine emails.

The study conducted by Welk et al. confirm the results of the study by Butavicius et al. related to impulsivity [106]. Welk et al. aimed to determine how individual differences relate to performance on a phishing task by having undergraduate students complete

a questionnaire and an email task in which they had to discriminate between legitimate emails and phishing attempts. The results indicated that certain trust, personality and impulsivity predictors were linked with accuracy in terms of detecting phishing attempts: personality characteristics that support reserved behavior, low impulsivity and distrust decreased phishing susceptibility in an email-based decision-making task.

Papers reporting no impact [99], Zielinska et al. conducted a questionnaire-based study with 96 participants recruited from Amazon Mechanical Turk. The study did not find any differences in age or gender in terms of susceptibility to phishing.

Similar findings were obtained in a study conducted by Kumaraguru et al. [21] that analyzed data from 42 participants and a study by Mohebzada et al. [48] involving 10,000 participants. However, instead of age and gender, [48] report that awareness is a critical parameter, as 10% of the users investigated in their study fell victim to phishing. Benenson et al. investigated whether there is a statistical correlation between the following factors and the click rate on phishing links:

- Gender of the sender and receiver
- Subjects are friends on Facebook
- Sender has a publicly available Facebook profile
- Receiver has knowledge that emails can be spoofed
- Receiver knows that clicking on links in emails can be dangerous

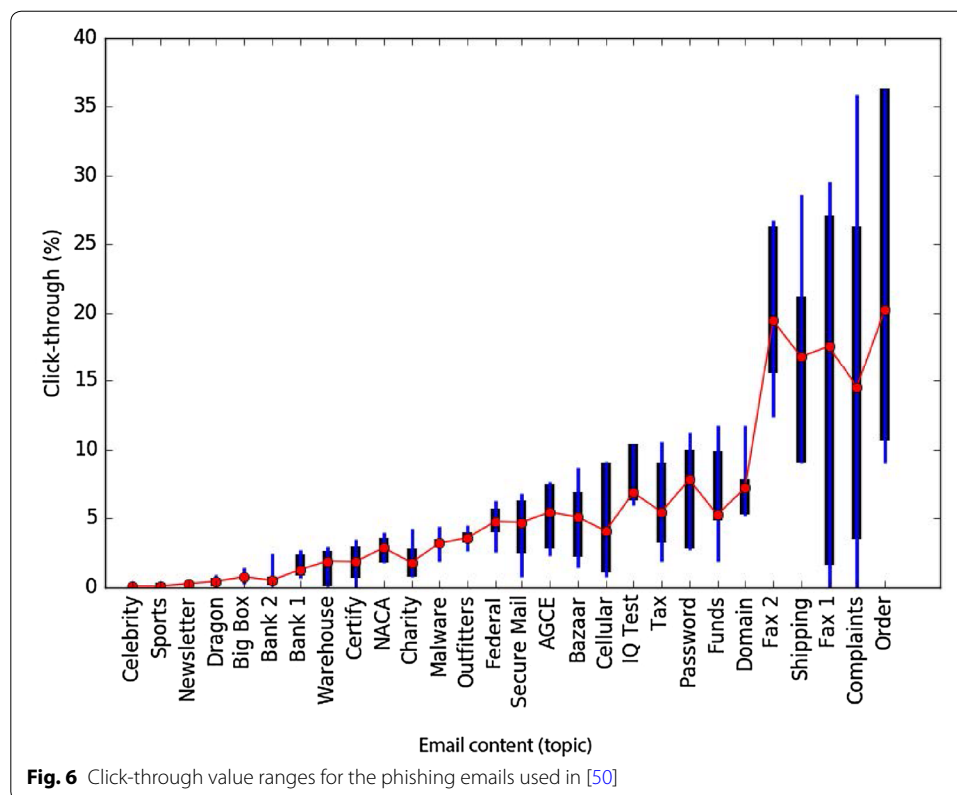
Their results show no significant statistical correlation between these factors and the clicking behavior [54]. Karumbaiah et al. [25] found the same for gender but also the personal traits of trust and perceived internet risk.

Another adverse finding concerning user properties has been reported by Leukfeldt et al. [49]. Their study shows that frequently engaging in online activities such as participating in chat rooms, gaming, actively using forums or engaging in high-visibility social networking is not correlated with an individual's susceptibility to phishing; furthermore, operating system or browser affinity was not found to be correlated with their phishing susceptibility as well.

The last no-impact findings concern the parameter of the technical complexity of an individual's job. In [82], Kumaraguru et al. find that employees working in technical and non-technical jobs exhibit similar susceptibility to phishing. This is confirmed by [81], who states that even educated users can fall victim to phishing as their detection ability alone may not be enough to prevent an attack. The authors also argue that contextual factors indirectly influence phishing susceptibility. They conclude that individuals fall victim to phishing attempts due to their lack of cognitive involvement rather than an inability to detect phishing.

Summary

Table 5 summarizes what the surveyed body of work reports regarding the impact or lack thereof user-specific properties. One key observation is that, with the exceptions of the properties age, gender, frequent engagement of online activities, and job technicality, the answer to the question of whether or not a parameter has an impact on susceptibility to phishing attacks seems quite clear. However, especially for properties that were



discussed by only one of the publications, the observation should be taken with a grain of salt.

As many parameters have been identified as impacting susceptibility to phishing attacks and given that there may be many more, an efficient approach could be to start training all employees using the same framework. In a subsequent step, a training regime (i.e., differentiation) could then be developed based on their response to training and progress using models such as the SCAM or the CRI proposed by Vishwanath et al. [80].

Email content and structure

Introduction

This section covers essential aspects one should consider when designing and populating a phishing email to use in anti-phishing training exercises. Such aspects could be the email's visual appearance, how the link Uniform Resource Locator (URL) is masked, or the content's context (see Fig. 7). Multiple studies have investigated how these properties influence the success rate of phishing emails. Analyzing these results enables the creation of synthetic phishing emails with varying levels of difficulty in terms of detection.

Survey

Siadati et al. [50] conducted a study on how the content of a phishing email impacts its success. They investigated which topics were more appealing to the participants in their study, as well as whether more persuasive content influences the outcome of phishing attempts. The results clearly show that persuasive emails do, indeed increase the

success rate. Moreover, the following five email topics were identified as the most effective (see Fig. 6 for more details): shipping, order, received fax email template #1, received fax email template #2 and complaints. In Fig. 6, the thinner bar shows the click-through rates of individual groups, while the thicker bar shows the range for a click-through rate of individual campaigns. Red dots show the weighted average click-throughs over the campaigns. On the other end of the spectrum, topics such as celebrity, sports, or newsletter are the least ineffective ones yielding meager click-through rates. The top-five most effective topics have more than twice the click-through rates than the immediately following topics. Caputo et al. observed no statistically significant difference in the clicking rate when exposing participants to email contents that the authors divided into the following categories: *other gain*, *other loss*, *individual gain* and *individual loss* [55]. For example, an email from the category *other loss* would state that another individual would suffer a financial loss should the recipient not click on the link provided.

Harrison et al. [87] studied how perceptions of social presence in a phishing attack influence the victimization rate. In their experiment, their participants were subject to a simulated phishing attack in which the amount of social presence in the email used was varied. Their results show that richness cues in the email were heuristically rather than systematically processed and that these cues significantly increased the likelihood of successful victimization. The authors, therefore, conclude that the rich information in phishing emails triggered perceptions of social presence and that the resulting heuristic evaluation increased the chances of victimization. Additionally, it appeared that once triggered, the perceived social presence of a phishing email not only reduced the users' considerations of mediation but also indirectly increased the persuasiveness of the email.

In [136], Parsons et al. concluded that the participants in their study developed personal approaches to the categorization of emails. They tended to treat emails as if they were important, regardless of their actual legitimacy. For instance, emails from banks or government institutions were more likely to be considered as important and therefore treated as legitimate. Additionally, the authors conclude that the participants were more likely to fall victim to phishing emails if their content threatened a potential financial loss on the part of the receiver. A similar result was presented by Butavicius et al. [127], who found that the most effective social engineering strategy for influencing a user's judgment of a link was authority, while the least effective was social proof. Their participants were unable to reliably distinguish between spear phishing and legitimate emails when the emails contained a reference to an authority figure. Thus, the authors concluded that, in terms of judging an email's legitimacy, the link destinations were unrelated to the actual content of an email. The study conducted by Jansen et al. investigated judgmental heuristics employed by users in evaluating the authenticity of messages [129]. Their participants' opinions about the validity of a website relied heavily on the presence of safety signs, such as a closed padlock symbol (presumably, however, they were unaware of how easily such a symbol can be faked). The study conducted by Dhamija et al. confirms this behavior [111].

In [130], Parsons et al. attempted to determine the best cues for identifying phishing emails and whether users actually use them. The authors surveyed studies related to this question, compiled a list of cues identified therein, and organized their findings

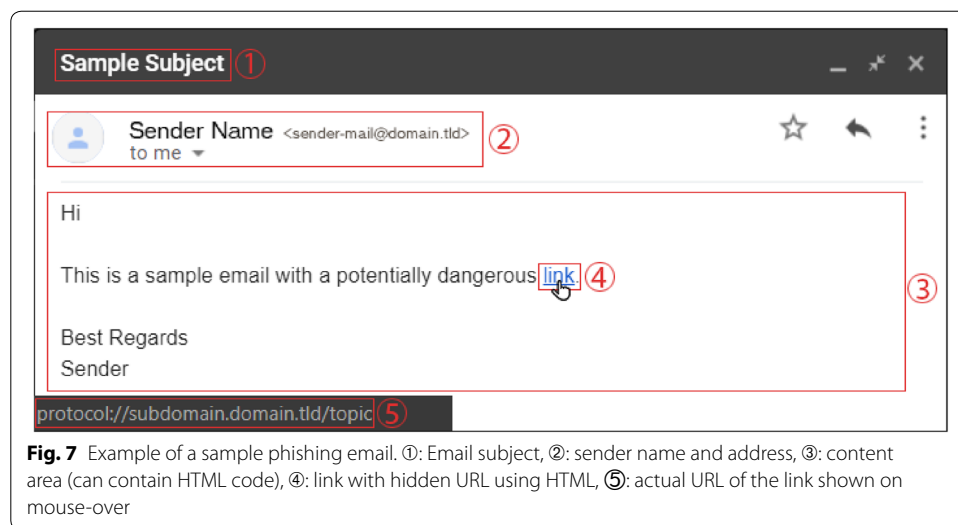
Table 6 Email cues identified in previous research by Parsons et al. [130]

Cue	Description
Consistency	Structure and focus of information
Links	URL, https, address bar
Visual presentation	Logos, banners, visual presentation, general design and look
Personalization	Personalization of the content inc. language and content aspects
Security	Security indicators and status bar
Spelling and grammatical errors	Spelling and grammar (mistakes)
Legal	Copyright information and legal disclaimers
Sender	Sender, his or her address, contact methods
Familiarity	Credibility and level of trust in source
Importance	Rational appeals
Urgency	Time pressure, overly urgent or forceful language/content
Positive and negative consequences	Emotional and motivational appeals, premise of the appeal, underlying motive of the website (and potential incentives as positive consequence)

into categories, as shown in Table 6. Because they found that all of the studies investigated were based on participants self-reporting how they used these cues to distinguish between phishing and genuine emails, the authors performed experiments to measure the impact thereof empirically. They identified content consistency, link legitimacy, email personalization, and spelling as the best indicators. However, their results indicate that users often make their decisions based on poor indicators; for example, their participants were influenced by the visual presentation of the email used. If the phishing email was visually more appealing (e.g., a professional-looking logo was present), they tended to make more accurate decisions concerning its legitimacy compared to emails with a poor visual presentation. Additionally, the authors found that participants were influenced by the urgent tone of an email, as they seemed to perform the worst in that case.

Similarly, Benenson et al. [32] studied why users click on the links provided in phishing emails. The results indicate the following reasons: 34% of users stated that they opened emails due to curiosity concerning their content—For example, the content may have been related to the actual behavior or activities of the recipient, such as a link to photographs of a party. 27% of users opened emails to determine their validity. 17% of users opened as they claimed to know the sender of the email, even though the addresses were generated with a random name selector. 16% of the participants opened because they trusted the technical solutions in place to keep them safe. Figure 3 shows a tricky combination of *content fits actual behavior or activities* and *curiosity*.

The study [79], which focused on students, found an increased phishing success rate when emails that are as similar as possible to the original were used. Additionally, more users fell victim to phishing when the linked page was an identical clone of the expected original website. According to Afroz et al. [124], most users will consider a website and will provide the requested information, if what they see does not contradict their expectations. The authors' analysis revealed that over 90% of users use a website's appearance as an indication of its authenticity. The goal of an attacker would, therefore, be to design a phishing website in such a way that it is as close in appearance to the original as possible.

**Table 7** URL spoofing tricks categorized as described by Canova et al. [128]

Category	Description	Example
1	Internet Protocol (IP) address as URL, no brand	http://130.82.162.6
2	Random/unrelated/trustworthy URL. Does not contain the company name.	http://account.com
3	Random/unrelated/trustworthy URL. Company name at the place of the department.	http://paypal.account.com
4	Random/unrelated/trustworthy/IP domain. Company name at the place of the topic/path.	http://account.com/www.paypal.com
5	Derived domains: The who section seems similar to the the real URL but uses an additional term	http://facebook-login.com
6	URL that contains well-hidden typos	http://www.twitterter.com
7	URL in which chars are replaced by similar-looking letters and numbers	http://www.arnazon.com

There are also works focusing on the features of URLs embedded in phishing emails (see Fig. 7—items ④ and ⑤). Canova et al. [128] defined multiple categories, each of which includes several URL spoofing tricks; these are listed in Table 7. Their results indicate that URL categories 1, 2, and 7 were the easiest to identify, where types 5 and 6 were the most difficult to spot and, therefore, the most successful. Subsequently, they published a follow-up study adding the results of a retention test conducted 5 months after the initial training. The attack using well-hidden typos (category 6) was again the most successful, where over 60% of the participants were unable to identify the message as phishing. Furthermore, the authors report that including keywords such as “secure” in an URL and sub-domain tricks (see category 3) confused the participants the most [128]. Andric et al. [79] found that users demonstrated superior performance in terms of identifying phishing URLs and fake websites when they knew the correct URL and the protocol used by the original website.

There are also contradictory results in the literature concerning the effect of email content and structure. Harrison et al. [84] designed multiple phishing emails, to which they added typographical/spelling errors. They found that all their efforts went

Table 8 Email content and structure parameters and their impact on the success rate of phishing messages

Parameter	Has impact	No impact
Content	Persuasive [50, 96], trust symbols [111, 120, 129], spelling [130], links [130], content consistency [130], personalization [130], visual appearance [95, 119, 130], urgency [130], social presence [87]	Spelling errors [84], link destination [127]
Topic	Shipping [50], order [50], received fax [50], complaint [50], banks [136], government institutions [136]	Other gain [55], other loss [55], individual gain [55], individual loss [55]
Link URL	Categories 5, 6 [128], same protocol [79], contains secure or similar terms [128]	Categories 1, 2, 7 [128]
Design	Clone of original [79, 124]	–

completely unnoticed and subsequently did not affect either processing or susceptibility to phishing.

Summary

In order to maximize the effectiveness of a phishing email, we could use a combination of the previously described study approaches. For example, the use of an extremely persuasive topic such as *shipping* or *order*, an email that looks identical to a regular email and an URL using spoofing tricks that fall into category 5 or 6 redirecting to a clone of the expected website is promising. Multiple studies found that the success rate of phishing attempts improves when emails that are very similar to the original one are used. However, the study of Harrison et al. [84] found that spelling errors have no impact. This may be because people do not spot them. According to Rawlinson et al. [140], the human brain can read words with scrambled characters because it generally processes word features through a classification/identification scheme. The brain can recognize a word as long as the beginning, and the end of the word remains intact, and the middle part of the word still contains the correct letter features, although they can be arranged independently of their correct position. This leads to the assumption that the visual appearance of an email is more important than the words used, with the limitation that the topic and content must still match that expected of an email. To provide a concise overview, Table 8 summarizes all of the findings described in this section.

Feedback

Introduction

This category covers the design of learning materials, when or how educational documents are presented to participants, and how a training program should be designed. Potential approaches to education could include courses, repeatedly sending educational material to target users, or attempting actually to phish users and presenting the relevant training material thereafter. The latter method is referred to as *embedded training* in this work. This section is organized as follows: First, we present results regarding how the training itself should occur (e.g., if courses are a more effective form of training than just providing informative material via email). In the second part, we analyze publications studying the training materials themselves (e.g., how the documents should be structured or whether more graphics should be used than text).

Survey

a. Form of the training: Based on their results, Kumaraguru et al. [82] suggest that users learn more effectively when the training materials are presented after the users have fallen victim to a simulated attack. The authors refer to this educational method as *embedded training*. Additionally, they also measured the average time the users spent reading the provided training materials. Participants in the embedded training group spent 97 s on average, whereas the non-embedded group spent 37 s. This result is reflected in data collected from recurring phishing tests: The adoption of the embedded approach results in an improved training effect. Al-Daeef et al. [23] also confirmed this finding by observing that users make better decisions concerning phishing emails after having to experience embedded training. Also, Kumaraguru et al. [21] did not observe a significant difference in phishing detection performance between the participants receiving non-embedded training and the members of the control group.

Offering personalized training is instrumental in increasing the effectiveness of the anti-phishing training program. The literature survey in [24] highlights the benefits of ongoing, embedded anti-phishing training for employees as such education will not be as detached from a user's reality as, for instance, a dedicated course would. Schroeder's suggestion is to implement training on a per-user basis with different difficulty levels. The author notes that incorporating personalized spaced repetition provides added benefits for employees, as they receive the impression that the training has been customized to their needs. The participants would feel more engaged by the customized materials since they knew that the training was designed to provide them the ability to succeed. Mapping the learning tasks to each level would allow each participant to progress at his or her own pace. Users might stretch themselves to reach a higher level than they would in the absence of a personalized program.

Carella et al. [22] confirmed that embedded training substantially outperforms no-training and in-class training situations. However, the authors stated that in-class training has the most significant short-term impact. The high short-term training effect of in-class education was also observed by Karumbaiah et al. [25], who, in their research, concluded that users exposed to a high-quality anti-phishing training video were less likely to click on phishing links during a subsequent 30-min experiment than those exposed to other training methods.

The phishing type against which users should be trained also impacts the effectiveness of embedded training. Caputo et al. [55] studied embedded training for spear phishing and obtained mixed results. They concluded that the training was not as effective against spear-phishing as it was against general phishing. The authors speculated that the participants might have perceived the provided information as "not credible, relevant or interesting".

How phishing education is presented to users has a significant impact on how they react to it. Wang et al. [102] extended an email client with a phishing warning bar, which would warn the user should he or she receive suspicious emails. However, the results showed that many users did not notice the warning sign and fell victim to the phishing attempt. Akhawe et al. [121] conducted a large-scale study to investigate the impact of warning messages further and found that such messages can indeed be effective in practice. The authors evaluated browser telemetry data obtained from Mozilla and Google

and reached the following conclusion: When malware or phishing warnings were shown, only a quarter of the users ignored the warnings and continued to open the website. If, however, the Secure Socket Layer (SSL) warning page was displayed, more than 70% of users clicked through. The authors concluded that the experience of a user for a specific warning message has a significant impact on the click-through rate. According to Engelman et al., such warning messages must be designed such that they actively interrupt the user's primary task, only pop up if necessary and require the user actually to read the message; besides, to be efficient, they should display clear and understandable choices [103].

b. Educational material: Kumaraguru et al. investigated whether users provided with text- or comic-based *training materials* exhibit different learning results in [75]. The participant group provided with the comic-based materials achieved better results than with standard training methods. The authors subsequently improved their training materials even further by developing a game called "Anti-Phishing Phil". Their results show that participants who played the game performed better in terms of identifying phishing URLs [57]. In a similar vein, Sheng et al. [30] studied and tested several anti-phishing materials, finding that there is no significant difference between the training effect of the materials as long as users are provided with at least one of them. A similar result was obtained by Jensen et al. [26], who concluded that training materials consisting of only text were as effective as those featuring a text-plus-graphics presentation method. Harrison et al. [84] suggest focusing the training on "refining the quality of initial attention to the email", such as by teaching users to focus on a few key elements of an email (e.g., the existence of hyperlinks or verifying the sender's email address). Greene et al. [101] analyzed the data of a 4.5-year-long embedded training-based phishing awareness program. They found that the people who clicked on the simulated phishing messages tend to overestimate the technological phishing detection system of their company. Therefore, they advise that companies should consider explicitly informing their employees that no technological solution is completely infallible. Promising training effects were identified in the results obtained by Siadati et al. [50], who developed a web-based interactive email client in which participants had to identify a certain number of suspicious elements to complete their training.

In their study, Kirlappos and Sasse [108] proposed that the way in which security education is designed should be revised. Their results show that materials provided to employees are largely ignored because they focus on indicators that users potentially do not understand or trust. Therefore, the authors propose offering different modules when implementing a training program, as they conclude that awareness, education, and training are three distinct steps in improving a user's security competence.

Zikai et al. show in [117] an interactive form of awareness training with a role-playing game. In their study, they compared their game to similar approaches and state that users learn the concepts of phishing better with playing their game than watching video material.

Summary

The effectiveness of anti-phishing training based on the embedded model has been successfully verified in the past. Ideally, such training should be designed on a

Table 9 Impact of the form of the training and the educational material

Parameter	More impact	Less/no impact
Form of anti-phishing training		
Short-term training	In-class [22, 25]	–
Long-term training	Recurring [24] embedded training [21–24]	–
Warning messages	Effective [121] if interrupting user [103]	Toolbar [102]
Educational material		
Form	Comic [26, 75], game [57, 117], text [26], does not matter [30]	Text [75]
Content	Key identifier [84], no technological solution is completely infallible [101]	–

per-participant basis as an ongoing process within an organization, starting with an in-class training seminar. Various types of training materials have been investigated, with mixed results being obtained (see Table 9), while one of the works considered that providing materials, regardless of their type, was the most important factor [30].

Knowledge retention

Introduction

This section covers works that investigate the impact of the anti-phishing training program over time. It presents findings related to the question of whether a single training session is sufficient or whether recurring training sessions at certain frequencies are required to achieve and maintain a decreased likelihood of employees falling victim to phishing attacks.

Survey

There are various findings that support the view that an effective anti-phishing training program should consist of multiple recurring training sessions [21, 24, 55, 75]. However, findings regarding how long participants retain the knowledge obtained during training or how long the intervals before potential re-training sessions should be, differ. On three occasions, Kumaraguru et al. concluded that users can retain learned content for at least 1 week [57, 82, 109]. On a similar time-scale, Jackson et al. [110] showed that users retained their anti-phishing knowledge up to 16 days after undergoing their first training session. Another study published by these authors titled “School of Phish: A Real-World Evaluation of Anti-Phishing Training”, confirmed knowledge retention even after 28 days [75].

The considered studies overwhelmingly conclude that training should be designed as an ongoing and integrated process. Employees should be able to train in a way that feels natural for them; for example, training could be integrated into their routine work activities. It was found in [23, 24] that, through ongoing anti-phishing training, click rates were reduced from 58 percent to single-digit percentages after the first training iteration. In [24], Schroeder further advises choosing the training intervals on an individual basis per user depending on his or her educational progress. These intervals should, however, be determined in such a way that they do not annoy employees by resulting in excessively frequent scheduled training sessions; however, each user should be trained at

Table 10 Knowledge retention summary

Parameter	Value
Approach	General recurring training [21, 24, 55, 75, 109], recurring training individualized per user [24]
Minimum interval	Seven days [57, 82], 16 days [110], 28 days [75]
Maximum interval	Set by management [24], less than five months [128]

Table 11 Proposed anti-phishing training program parameters

Parameter	Section	Value(s)
Susceptibility	Target group impact	Train everyone
User specific training	Target group impact	Use a model similar to Cyber Risk Index (CRI) to identify the appropriate training method
Email design	Email content and structure	1:1 clone of the legitimate mail
Best email topics	Email content and structure	Shipping, orders, received fax
Email persuasiveness	Email content and structure	More = better
Education progression	Email content and structure , Feedback	Level system, per user
Level design	Email content and structure , Feedback	Increasing difficulty (see " Email content and structure " section)
Education form	Feedback	Initial course then ongoing training based on a user's weaknesses as identified by the CRI
Feedback	Feedback	Embedded training, imminent
Training interval(s)	Knowledge retention	Adjusted to levels, min. 4 x/year

least four times a year. In a similar vein, Canova et al. [128] found a significant decrease in the performance of the participants in their retention study after 5 months.

Summary

As summarized in Table 10, all of the works considered in this survey agree on the notion that recurring training sessions must be scheduled to ensure that the learned anti-phishing knowledge is not forgotten. Unfortunately, the findings regarding knowledge retention are not as clear. They suggest that the retention period is between 7 days and 5 months. Therefore, one should train all users at least once every 5 months even with an optimistic view on knowledge retention.

Discussion

Our literature analysis showed that anti-phishing training has a significant impact on user susceptibility to phishing attacks. It is, therefore, evident that any organization should have a valid and well-founded anti-phishing training program. However, a key question lingers: what should such a program look like?

Our comparative analysis of related works showed that the parameters and values listed in Table 11 are reasonably certain to have a positive impact when they are taken into consideration in a program's design.

Based on these parameters, we first discuss how such training should look alike. After that, we consider the implications for tools that can be used to implement or facilitate

anti-phishing training. Finally, we conclude this section with a brief look at the current state of anti-phishing training tools.

Aspects of a well-founded anti-phishing training program

The reviewed body of work leaves little doubt that everyone is susceptible to phishing, to at least some degree. Therefore, every organization should have an anti-phishing training program. A valid and well-founded anti-phishing training program should start with a “kick-off course,” as training sessions organized as courses produce the highest short-term training effect. After this initial step, the participants of the program should be trained through *embedded training*. However, the reviewed literature suggests (see “[Target group impact](#)” section) that user-specific parameters such as age, gender, technical expertise, and personal traits have an impact on phishing susceptibility and on the type of training that yields the best results. A method such as the CRI could be used to determine how much and what type of training an employee needs.

Concerning embedded training, the training material must be displayed as soon as a mistake is made; for example, just after clicking a link in a phishing email. Alternatively, the presentation of the training material could be delayed until some additional steps are taken, for example, after credentials have been entered on a fake company login page. However, these cases are challenging, as, if the user clicks but does not enter his/her credentials, a training action might still be required if the phishing attempt could have been recognized based on the email content and link. The training material itself should provide information on why the user is being presented with that and how he or she could have recognized this instance of a phishing attempt. Moreover, if an employee does not click on the phishing email but does not report it either, he/she should receive training materials on how to report phishing emails and why this step is essential.

As shown in Table 11, according to the research findings in the literature, the email topics leading to the highest click rates are *shipping*, *orders* and *received fax*. However, this situation does not mean that only such emails should be used. According to the context and training goals, it is also appropriate to use other topics or a mixture thereof, with those that have a higher impact being weighted more than others (see Table 8).

Each employee will most certainly have a different knowledge state before training. A possible solution for this problem could be to create multiple difficulty levels and allow users to progress through those individually. These levels would contain different sets of emails and landing pages, with their difficulties being adjusted based on the findings presented in “[Email content and structure](#)” section. If employees continuously exhibit the correct anti-phishing behavior, we could upgrade them directly to the next level. Alternatively, we could send them an email describing their success and offering the possibility of proceeding to the next level if desired.

To ensure that the gained anti-phishing knowledge is retained, a program should be designed as an ongoing process that is integrated into users’ daily workflow and mimics actual attacks as closely as possible. Each user should be exposed to such training at least once every 5 months, but preferably four times a year. However, as the results regarding the ideal intervals between re-training sessions differ, one could experiment as suggested by Schroeder et al. [24]: The intervals should be chosen such that the users do

not get annoyed with excessively frequent retesting but still fulfill the requirements set by management.

Organizational aspect, perception and ethics

Although our conclusions may give the impression that intensive anti-phishing training should be implemented in any organization, there is also the organizational aspect of anti-phishing training regarding how such training might be perceived/taken up in different organizations.

Since organizations differ in their settings, security, and organizational cultures, the impact of a well-founded training programme may vary across companies. If, for instance, a company with a flat organizational structure and a very liberal work culture implemented embedded anti-phishing training, its employees could perceive the training as an observation tool. In contrast, in a bank, where each action taken by an employee might already be monitored, the likelihood of such impressions might be lower.

In any case, embedded training may increase the pressure on employees as, for various reasons, they may not wish to fail the training procedure. Therefore, they might feel constantly tested or pressured by their employers, which could have an impact on their health and/or work performance. Thus, any security training, including phishing training, should be varied according to the needs, market pressures, modernization goals, prerequisites, and budget of a firm.

Additionally, it is crucial to consider the “security fatigue effect”, referring to the situation in which people (e.g., a company’s employees) become overwhelmed by and tired of the barrage of installed security warnings and regulations [141]. They are basically “drowned” in the ongoing flow of advice concerning how they should stay safe and keep constantly alert. With regards to anti-phishing training, each company has to consider for itself whether its employees are able to handle additional training or whether they would become overwhelmed. If, for example, a company already has security training programs in place, adding an anti-phishing training program might prove counterproductive in terms of security fatigue, and the desired security improvement effect might therefore not materialize.

Psychological aspects of training design

The psychology of end-users, the subjects to be trained to achieve higher phishing-awareness, and attackers should be considered when devising effective anti-phishing training programs [142, 143].

Regarding the end-users, the literature review in “[Literature analysis](#)” section provides ample evidence that the exploitation of peculiarities of the human psyche is an important factor for a successful phishing attack. Examples of such peculiarities are the almost blind trust in authority figures [127] and security symbols [129] or the lower attention to phishing in emails with an urgent tone [130]. One way to address this is to design training programs that do not only teach users how to recognize phishing emails but aim at altering the user’s problematic cognitive processes. Other end-user related psychological aspects are the different perception of training if it is detached from the user’s reality [24] or the personal relevance of the training material. The former can be addressed with

embedded training and the latter by adjusting the training method and material to different psychological profiles. However, the reviewed literature does not offer much guidance on how these could be done.

Regarding the attackers, a better understanding of the psychology of attackers is crucial in modeling adversary behavior and identifying the implicit factors that determine how deception and phishing strategies are employed in phishing emails in the first place [74, 144]. Currently, the psychology of criminal behaviors is usually neglected in the field of cybersecurity [143]. Nevertheless, target-adversary interactions and how they are driven as part of adversary strategies are important in facilitating realistic phishing simulations and, consequently, training tools. Therefore, researching the application and exploitation of psychological aspects in the design and development of anti-phishing training seems promising.

Implications for anti-phishing training tools

To achieve the training objectives described above, the proposed parameters in Table 11 should be reflected in the design and capabilities of anti-phishing training tools. In addition, those tools should entail practical functions which enable the exploitation of these findings. More specifically, such tools should consider the following aspects:

Progression system: As each user must have the ability to progress at his or her own pace, the tool has to support some level-based progression system. The user is to be moved to a higher level automatically as soon as he or she has mastered the current one. Similarly, he or she could be placed at a lower level should he or she continuously fail to identify attacks at the present level of difficulty. For initial level positioning and training selection, the system could implement the CRI survey questions [80] and apply the algorithm developed by Vishwanath et al. to identify the appropriate training focus. Furthermore, to identify personality traits and cognitive processes that potentially amplify a user's susceptibility to phishing, a generic psychological questionnaire could be used. The level and progression of such users could then be adjusted according to the impact severity of their traits.

Emails: As users should be exposed to simulated phishing attacks of different difficulties, email templates must be assignable to levels, or the tool should have a mechanism by which to automatically adjust the given template to a specific level. Such adjustments might include changing the URLs used, adding safety signs to the content like a closed padlock [129] or tweaking the other parameters described in "Email content and structure" section. Another desirable automation is to mutate the URLs used in a campaign to reflect different URL categories (see Table 6). To minimize the administrative efforts required, the system should be able to autonomously manage the URLs on the company's internal Domain Name System (DNS) server.

To further reduce the work time that an administrator has to devote to creating templates, a tool could additionally offer community template pools that enable users to exchange templates. This solution would allow it to exploit the power of numbers (i.e., to offer support mechanisms by which to share the workload among the members of a community). Another approach to improve efficiency is to share the load to create email templates for typical user groups in the academic community. However, a major challenge is to determine who should operate and manage this platform, as well as how

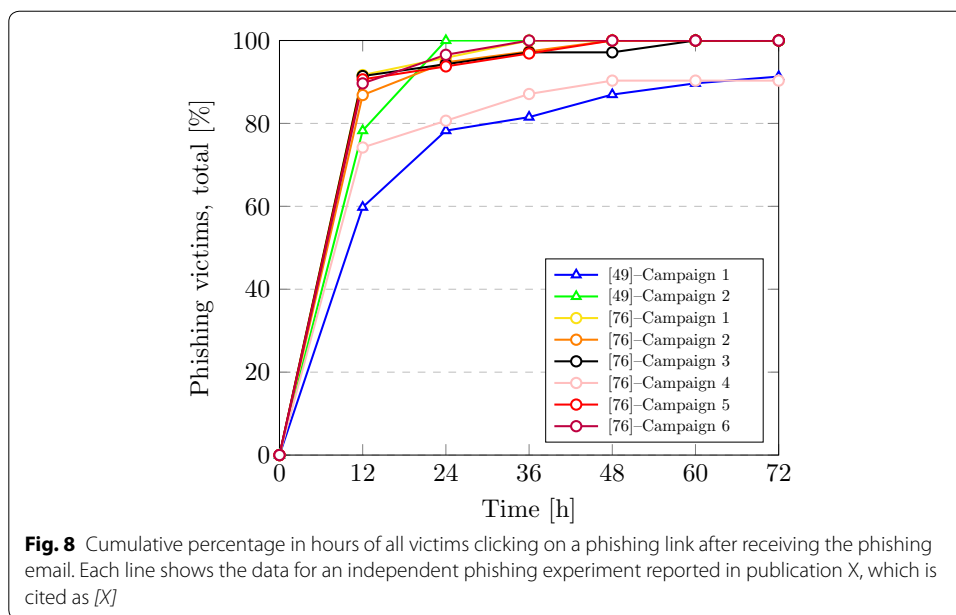
quality control for such a collaborative infrastructure should be implemented. Similar to the personality trait-based adjustments made in other parts of the training design, the emails should be adjusted based on the user's specific traits. For example, people who tend to overlook details can be trained with emails that are only slightly different from genuine messages.

Automatic population of user database: The system should feature an active connection to the central employee database of the company. This allows the automatic population of users during deployment or in the event of changes due to new requirements or role changes. The automatic inclusion of new employees is a particularly important requirement in terms of addressing vulnerabilities as fast as possible.

Feedback pages: The feedback page has to be adjusted according to the level a user is currently on. This means that it has to match the difficulty level of the simulated phishing email sent to that user. For example, on easier levels, it should directly show the user how he or she could have identified the phishing email. For advanced levels, first, a fake website with a login form could be displayed; then, if the user enters his or her login details, the feedback page with the training material could be displayed. If information about the personality traits of the users is available, the system could display feedback pages that are more suitable for a specific user's cognitive processing and the ability to assimilate information. For instance, if a user reacts better to graphical content, comic-based feedback pages could be shown.

Retraining system: To support the knowledge retention of employees, the tool should automatically schedule retraining sessions for all users. This should be done in a way that depends upon the user's current level, and it should feature some time randomization to prevent, for example, having all the phishing emails sent at 8 a.m. Additionally, similar to the *progression system*, if a generic psychological questionnaire is administered before the start of the training, the retraining-scheduling algorithm could take the user's psychological and cognitive traits into account as well. For instance, the retraining interval could be longer for users who exhibit traits that result in lower susceptibility.

Phishing email reporting system: If users spot a simulated phishing email in their inboxes, they must be provided with the ability to report this email as an instance of phishing. To support the principle of embedded training, which proposes training users in their working environments, this should be the same mechanism employees use to report real phishing threats. One such mechanism is forwarding the email to a special company email address; another is clicking on a "report email as phishing" button in the email client. Since offering the latter has become quite common in many email clients (e.g., the "Report phishing" drop-down menu option in Gmail, the "Report Message" add-in button for Microsoft Outlook or the "Report Spam" add-on for Thunderbird), the effort required to implement a reporting mechanism and integrate it into the daily routine of a company or its employees is probably acceptable. If a user neither falls for the phishing email nor reports it within the first 24 h, it is likely that it was missed, ignored, or not processed by the user, due to reasons such as being out of the office or having a day off. The period of 24 hours has been reported by [75] and confirmed by the observation made by Mohebzada et al. that many users fell for their phishing attempt even though their campaign was only active for 18 h. The results of [48, 75] are summarized in Fig. 8.



However, one challenge related to such reporting systems is analyzing the reported emails and providing feedback. If a company does not have the staff and processes in place to react to such reports and provide feedback, employees might lose interest in reporting phishing emails [138]. This lack of interest would, in turn, make measuring the impact of training efforts difficult. While there is some practical advice available from renowned sources (e.g., from the SANS Institute [145]) on how to design a good reporting process, it remains unclear which advice is backed by science and which is not. A literature review with a focus on the design of such a reporting system would be needed to shed more light on this question.

Privacy: Another requirement is the support of adequate privacy features. Although they are not related to the core performance of an anti-phishing training effort, such features are crucial for any practical tool to protect its user's privacy. To this end, statistics and tracking mechanisms should work with pseudonyms. Structural measures such as the isolation of analytics from the sending component in these tools are also necessary. Different aggregation and anonymization schemes for creating reports, such as k-anonymity and differential privacy, should be integrated into the tool [146].

Automated optimization of training parameters: Using data from multiple institutions, companies and sectors may also provide opportunities for synergistic gains, as analyzing the impact of different factors on training effectiveness could be made more streamlined and generic. Data sharing among stakeholders enables large-scale and long-term analysis with which the impact of different factors on training effectiveness can be measured. The results could then be used to automatically and continuously fine-tune the training parameters of the participating stakeholders.

Anti-phishing training tools and available features

There are intrinsic links among the factors that determine the success of a phishing attack, the effectiveness of anti-phishing training and the construction and operation

of an anti-phishing training tool. To get an idea whether or not today's anti-phishing training tools come with the functionality required to implement a training as outlined in the previous two subsections, we searched the Internet for such tools. More specifically, for each desired functionality or feature (e.g., a level-based approach to training where the training level is automatically determined and adjusted based on the user's feedback), we attempted to identify at least one commercial or non-commercial tool that offered it. In summary, we found that the currently available tools lack at least one of the desired features. However, since our findings are based on information that could be found using search engines and/or by browsing the respective webpages of each product only, we might have overlooked tools for which this information is not available through these channels. A more detailed summary of our most important finding regarding the aspects discussed in the previous subsection can be found below.

Progression system: We could not identify any tool that supports an individualized automated user progress tracking and level system. Automated tracking and modification of training intensity based on user feedback, personality traits, psychological processes, and progress have not been implemented in any anti-phishing training tool. Furthermore, scientific algorithms that could help to select targets and/or determine why an employee fails to identify phishing threats (e.g., the CRI) have not been implemented.

Emails: While most of the available tools support template mechanisms, they all lack the ability to categorize templates based on detection difficulty. Additionally, none offers a mechanism that can automatically alter a template to increase or reduce its detection difficulty. The shortcomings of existing tools include the lack of functionality by which to manage the URLs automatically used in training emails and mutations thereof in a company's DNS infrastructure. Some tools ship with templates mimicking emails of well-known Internet companies, such as Google or Amazon. Others offer a version of a template exchange platform based on Github, but we did not identify a tool with a directly-integrated platform usable for everyone.

Automatic population of user database: Most available tools support manual user imports, for instance via comma-separated values (csv) files or through a Lightweight Directory Access Protocol (LDAP) connection. These features must be extended to automatically pull new users from an organization's central user database and start their training.

Feedback pages: Most tools offer the ability to upload HTML content that will be presented to the user when he or she clicks on the link in the phishing email. This mechanism could be used to upload educational material. However, as most tools are using campaigns to send out phishing emails, such a page can only be defined on the campaign level. An ideal solution would require individual landing pages matching the sent phishing email and therefore matching the level a user is currently on; in addition, these landing pages should be displayed in a form that is adjusted to a user's psychological and personality traits. We could not find any available tool that offered such a feature.

Retraining system: As most tools use a campaign system, retraining cannot be applied as proposed in our analysis. After each campaign, an administrator would have to analyze the results of each user and manually schedule the follow-up training sessions.

Phishing email reporting system: Some of the tools we looked at offer a way for users to report an email as an instance of phishing. However, this mechanism is usually connected to the anti-phishing training system only and does not relay information about real threats to the company's email filtering solution.

Automated optimization of training parameters: Most tools assess and create reports on the performance of users and create a report about it. However, we could not identify any tool that continuously analyzes the impact of training parameters, such as training frequency or email types, on training effectiveness. Therefore, we could not find any available tool which makes recommendations on how to modify the training parameters for enhancing the training gains.

Conclusion

Phishing is a growing security issue for both institutions and individuals. Although there are various mitigation techniques, proactive anti-phishing training is an important building block of any multi-level phishing defense strategy. In this paper, we identified various factors that influence the effectiveness of such training efforts. Building on our analysis of the research literature, we outlined how an effective anti-phishing training program should be designed and implemented. Based on the weak coherence between our empirical findings and currently used anti-phishing training solutions, we believe that this contribution addresses a crucial technical gap.

In our discussion, we outlined several implications of our findings concerning the design and capabilities of anti-phishing tools. Significant design aspects and capabilities in this regard are automated operation and individualization with continuous assessment/optimization of the configuration of training parameters. This is crucial, as our literature analysis showed that research results concerning some of the parameters are inconclusive or even contradictory, indicating that these parameters require further investigation. Moreover, an effective anti-phishing training tool should have community functions to facilitate cooperation and load-balancing among disparate anti-phishing efforts (e.g., shared email templates or co-designed training curricula.)

Based on our survey and analysis of relevant sources in the technical literature, we found that, despite the various advanced capabilities that tools currently available in the anti-phishing domain offer, such tools only support a limited subset of the potential factors identified as necessary to yield the desired training effects. Therefore, we believe that our work does have a high practical value in terms of contributing to the development of more complete training solutions with a more significant impact to reduce phishing susceptibility on the part of users. We are convinced that greater awareness of phishing techniques and means of addressing them increases overall security and peace of mind.

Future research directions

Our survey points out that two key research directions: First, the factors on anti-phishing training effectiveness deserve further research with more extensive and diverse experiments in higher numbers focusing on the gray areas, i.e., where contradictory results are available in the current body of work. Second, phishing awareness training, as done today, has several limitations. First of all, this includes a lack of consideration of

scientific results that are available in the scientific outputs. Furthermore, a lack of individualization of training limits the efficiency of training. Thus, how to customize training based on trainee profiles is another research topic.

Once the training is done, another requirement becomes evident, leading to another research direction: how to measure the training effect. This is accompanied by the lack of ensuring long-term training benefits. More studies are necessary to show the long-term effects of anti-phishing training and make the results of these studies comparable. To this end, we will conduct further iterations of our survey in the future to see how the research in this field has progressed over the years and to compare new findings.

Acknowledgements

Not applicable.

Authors' contributions

DJ and BT did the initial survey design. DJ carried out most of the survey work with TS, GG and BT helping out at a later stage of the survey work. GG and BT supervised the drafting of the manuscript and contributed heavily to its design and revisions. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 3 December 2019 Accepted: 15 June 2020

Published online: 09 August 2020

References

1. Infosec: phishing definition, prevention, and examples (2019). <https://resources.infosecinstitute.com/category/enterprise/phishing/>
2. Bissell K, LaSalle RM, Cin PD (2019) Accenture's ninth annual cost of cybercrime study: unlocking the value of improved cybersecurity protection. <https://www.accenture.com/us-en/insights/security/cost-cybercrime-study>
3. Nero PJ, Wardman B, Copes H, Warner G (2011) Phishing: crime that pays. In: 2011 eCrime researchers summit, pp 1–10
4. Bisson D (2015) Sony hackers used phishing emails to breach company networks. <https://www.tripwire.com/state-of-security/latest-security-news/sony-hackers-used-phishing-emails-to-breach-company-networks/>. Accessed 26 Dec 2017
5. Sanger DE, Benner K (2018) U.S. accuses North Korea of plot to hurt economy as spy is charged in Sony hack. The New York Times, Chap, U.S. Accessed 29 Oct 2018
6. Franceschi-Bicchieri L (2016) Russian hackers launch targeted cyberattacks hours after trump's win. https://motherboard.vice.com/en_us/article/nz79gb/russian-hackers-launch-targeted-cyberattacks-hours-after-trump-s-win. Accessed 26 Dec 2017
7. Aaron G (2020) APWG phishing activity trends 4th quarter report 2019. https://docs.apwg.org/reports/apwg_trends_report_q4_2019.pdf. Accessed 04 Jan 2020
8. Aaron G (2019) APWG phishing activity trends 4th quarter report 2018. https://docs.apwg.org/reports/apwg_trends_report_q4_2018.pdf. Accessed 04 Jan 2020
9. Aaron G (2018) APWG phishing activity trends 4th quarter report 2017. https://docs.apwg.org/reports/apwg_trends_report_q4_2017.pdf. Accessed 04 Jan 2020
10. Aaron G (2017) APWG phishing activity trends 4th quarter report 2016. https://docs.apwg.org/reports/apwg_trends_report_q4_2016.pdf. Accessed 04 Jan 2020
11. Hong J (2012) The state of phishing attacks. *Commun ACM* 55(1):74–81
12. Gorman S (2013) Annual U.S. cybercrime costs estimated at \$100 billion. *Wall Street J.* Accessed 22 Mar 2017
13. Morrow S (2019) Juniper research—the future of cybercrime & security research report. <https://www.juniperresearch.com/document-library/white-papers/the-future-of-cybercrime-white-paper>
14. Cybersecurity ventures: 2019 official annual cybercrime report (2019). <https://www.herjavecgroup.com/the-2019-official-annual-cybercrime-report/>
15. CNBC: Xoom says \$30.8 mln transferred fraudulently to overseas accounts (2015). <https://www.cnbc.com/2015/01/06/xoom-says-308-mln-transferred-fraudulently-to-overseas-accounts.html>
16. Dou Z, Khalil I, Khreishah A, Al-Fuqaha A, Guizani M (2017) Systematization of knowledge (SoK): a systematic review of software-based web phishing detection. *IEEE Commun Surv Tutor* 19(4):2797–2819
17. Gupta BB, Tewari A, Jain AK, Agrawal DP (2017) Fighting against phishing attacks: state of the art and future challenges. *Neural Comput Appl* 28(12):3629–3654

18. Deeb Al-Mo AA, Wan T-C, Tat-Chee K, Altaher A, Ramadass S, Manasrah A, Melhiml LB, Anbar M (2011) An online model on evolving phishing e-mail detection and classification method. *J Appl Sci* 11(18):3301–3307
19. Angelov P, Filev DP, Kasabov N (2010) *Evolving intelligent systems: methodology and applications*. Wiley, Hoboken
20. Fette I, Sadeh N, Tomasic A (2007) Learning to detect phishing emails. In: *Proceedings of the 16th international conference on world wide web. WWW '07*, ACM, New York, pp 649–656
21. Kumaraguru P, Rhee Y, Sheng S, Hasan S, Acquisti A, Cranor LF, Hong J (2007) Getting users to pay attention to anti-phishing education: evaluation of retention and transfer. In: *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit*, ACM, Pittsburgh, Pennsylvania, pp 70–81
22. Carella A, Kotsoev M, Truta TM (2017) Impact of security awareness training on phishing click-through rates. In: *2017 IEEE international conference on Big Data (Big Data)*, pp 4458–4466
23. Al-Daeef MM, Basir N, Hukins M (2017) Security awareness training: a review. In: *Proceedings of the world congress on engineering 2017*, vol 1
24. Schroeder J (2017) Persistent training. In: *Advanced persistent training*, Apress, Berkeley, pp 25–32
25. Karumbaiah S, Wright RT, Durcikova A, Jensen ML (2016) Phishing training: a preliminary look at the effects of different types of training. *WISP 2016 proceedings*. 11
26. Jensen ML, Dinger M, Wright RT, Thatcher JB (2017) Training to mitigate phishing attacks using mindfulness techniques. *J Manage Inf Syst* 34(2):597–626
27. SANS: SANS security awareness—phishing (2019). <https://www.sans.org/security-awareness-training/ouch-newsletter/2015/phishing>
28. MITRE: MITRE attack framework—initial access (2019). <https://attack.mitre.org/tactics/TA0001/>
29. Yue C, Wang H (2010) Bogusbiter: a transparent protection against phishing attacks. *ACM Trans Internet Technol* 10(2):6–1631
30. Sheng S, Holbrook M, Kumaraguru P, Cranor LF, Downs J (2010) Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp 373–382
31. Seals T (2017) Cost of user security training tops \$290K per year. *Infosecurity magazine*. <https://www.infosecurity-magazine.com/news/cost-of-user-security-training>. Accessed 15 Sept 2017
32. Benenson Z, Gassmann F, Landwirth R (2016) Exploiting curiosity and context: how to make people click on a dangerous link despite their security awareness. *BlackHat USA*
33. Stembert N, Padmos A, Bargh MS, Choenni S, Jansen F (2015) A study of preventing email (Spear) phishing by enabling human intelligence. In: *2015 European intelligence and security informatics conference*, pp 113–120
34. Vrbanić G, Fister I, Podgorelec V (2018) Swarm intelligence approaches for parameter setting of deep learning neural network: case study on phishing websites classification. In: *Proceedings of the 8th international conference on web intelligence, mining and semantics. Association for Computing Machinery, New York*. <https://doi.org/10.1145/3227609.3227655>
35. Tian K, Jan STK, Hu H, Yao D, Wang G (2018) Needle in a haystack: tracking down elite phishing domains in the wild. In: *Proceedings of the internet measurement conference 2018. IMC '18*, Association for Computing Machinery, New York, pp 429–442. <https://doi.org/10.1145/3278532.3278569>
36. Sirigineedi SS, Soni J, Upadhyay H (2020) Learning-based models to detect runtime phishing activities using urls. In: *Proceedings of the 2020 the 4th international conference on compute and data analysis. ICCDA 2020*, Association for Computing Machinery, New York, pp 102–106. <https://doi.org/10.1145/3388142.3388170>
37. Tyagi I, Shad J, Sharma S, Gaur S, Kaur G (2018) A novel machine learning approach to detect phishing websites. In: *2018 5th international conference on signal processing and integrated networks (SPIN)*, pp 425–430
38. Sahingoz OK, Buber E, Demir O, Diri B (2019) Machine learning based phishing detection from urls. *Expert Syst Appl* 117:345–357. <https://doi.org/10.1016/j.eswa.2018.09.029>
39. Bahnsen Alejandro C, Ivan Torroledo LDC, Villegas S (2018) Deepphish: simulating malicious ai. In: *2018 APWG symposium on electronic crime research (eCrime)*, pp 1–8
40. Pham C, Nguyen LAT, Tran NH, Huh E, Hong CS (2018) Phishing-aware: a neuro-fuzzy approach for anti-phishing on fog networks. *IEEE Trans Netw Serv Manage* 15(3):1076–1089
41. Mayhorn CB, Nyeste PG (2012) Training users to counteract phishing. *Work* 41(Supplement 1):3549–3552
42. Alnajim A, Munro M (2009) An approach to the implementation of the anti-phishing tool for phishing websites detection. In: *2009 international conference on intelligent networking and collaborative systems, IEEE*, pp 105–112
43. Liu D, Wang W, Wang Y, Tan Y (2019) Phishledger: a decentralized phishing data sharing mechanism. In: *Proceedings of the 2019 international electronics communication conference. IECC '19*. Association for Computing Machinery, New York, pp 84–89. <https://doi.org/10.1145/3343147.3343154>
44. Hutchings A, Clayton R, Anderson R (2016) Taking down websites to prevent crime. In: *2016 APWG symposium on electronic crime research (eCrime)*, pp 1–10
45. Whitman ME (2003) Enemy at the gate: threats to information security. *Commun ACM* 46(8):91–95
46. Google: about Google Scholar (2019). <https://scholar.google.ch/intl/en/scholar/about.html>. Accessed 24 Apr 2019
47. Harzing A-W (2019) Publish or Perish. <https://harzing.com/resources/publish-or-perish>. Accessed 24 Apr 2019
48. Mohebzada JG, Zarka AE, Bhojani AH, Darwish A (2012) Phishing in a university community: two large scale phishing experiments. In: *2012 international conference on innovations in information technology (IIT)*, pp 249–254
49. Leukfeldt E (2014) Phishing for suitable targets in the netherlands: routine activity theory and phishing victimization. *Cyberpsychol Behav Soc Netw* 17:551–555
50. Siadati H, Palka S, Siegel A, McCoy D (2017) Measuring the effectiveness of embedded phishing exercises. In: *10th USENIX workshop on cyber security experimentation and test (CSET 17)*. <https://www.usenix.org/node/205854>
51. Gordon WJ, Wright A, Aiyagari R, Corbo L, Glynn RJ, Kadakia J, Kufahl J, Mazzone C, Noga J, Parkulo M, Sanford B, Scheib P, Landman AB (2019) Assessment of employee susceptibility to phishing attacks at us health care institutions. *JAMA Netw Open* 2(3):190393–190393. <https://doi.org/10.1001/jamanetworkopen.2019.0393>

52. Taib R, Yu K, Berkovsky S, Wiggins M, Bayl-Smith P (2019) Social engineering and organisational dependencies in phishing attacks. In: Lamas D, Loizides F, Nacke L, Petrie H, Winckler M, Zaphiris P (eds) *Human–computer interaction—INTERACT 2019*. Springer, Cham, pp 564–584
53. Baillon A, de Bruin J, Emirmahmutoglu A, van de Veer E, van Dijk B (2019) Informing, simulating experience, or both: a field experiment on phishing risks. *PLoS ONE* 14(12):1–15. <https://doi.org/10.1371/journal.pone.0224216>
54. Benenson Z, Gassmann F, Landwirth R (2017) Unpacking spear phishing susceptibility. In: Brenner M, Rohloff K, Bonneau J, Miller A, Ryan PYA, Teague V, Bracciali A, Sala M, Pintore F, Jakobsson M (eds) *Financial cryptography and data security. Lecture notes in computer science*. Springer, Cham, pp 610–627
55. Caputo DD, Pfleeger SL, Freeman JD, Johnson ME (2014) Going spear phishing: exploring embedded training and awareness. *IEEE Secur Priv* 12(1):28–38
56. Jagatic TN, Johnson NA, Jakobsson M, Menczer F (2007) Social phishing. *Commun ACM* 50(10):94–100
57. Kumaraguru P, Sheng S, Acquisti A, Cranor LF, Hong J (2010) Teaching Johnny not to fall for phish. *ACM Trans Internet Technol* 10(2):7–1731
58. Dodge RC, Carver C, Ferguson AJ (2007) Phishing for user security awareness. *Comput Secur* 26(1):73–80
59. Li W, Lee J, Purl J, Greitzer F, Yousefi B, Laskey K (2020) Experimental investigation of demographic factors related to phishing susceptibility. In: *Hawaii international conference on system sciences*. <http://hdl.handle.net/10125/64015>. Accessed 01 Apr 2020
60. Burns AJ, Johnson ME, Caputo DD (2019) Spear phishing in a barrel: insights from a targeted phishing campaign. *J Organ Comput Electron Commer* 29(1):24–39. <https://doi.org/10.1080/10919392.2019.1552745>
61. Gordon WJ, Wright A, Glynn RJ, Kadakia J, Mazzone C, Leinbach E, Landman A (2019) Evaluation of a mandatory phishing training program for high-risk employees at a US healthcare system. *J Am Med Inform Assoc* 26(6):547–552
62. Steves, MP, Greene KK, Theofanos MF (2019) A phish scale: rating human phishing message detection difficulty. In: *Workshop on usable security (USEC)*
63. Ikhsan MG, Ramli K (2019) Measuring the information security awareness level of government employees through phishing assessment. In: *2019 34th international technical conference on circuits/systems, computers and communications (ITC-CSCC)*
64. Higashino M, Kawato T, Ohmori M, Kawamura T (2019) An anti-phishing training system for security awareness and education considering prevention of information leakage. In: *2019 5th international conference on information management (ICIM)*, pp 82–86
65. Rastenis J, Ramanauskaitė S, Janulevičius J, Čenys A (2019) Credulity to phishing attacks: Areal-world study of personnel with higher education. In: *2019 Open conference of electrical, electronic and information sciences (eStream)*
66. Diaz A, Sherman AT, Joshi A (2020) Phishing in an academic community: a study of user susceptibility and behavior. *Cryptologia* 44(1):53–67
67. Kim B, Lee D-Y, Kim B (2019) Deterrent effects of punishment and training on insider security threats: a field experiment on phishing attacks. *Behav Inf Technol* 0(0), 1–20
68. Canfield CI, Fischhoff B, Davis A (2019) Better beware: comparing metacognition for phishing and legitimate emails. *Metacogn Learn* 14(3):343–362
69. Xiong A, Proctor RW, Yang W, Li N (2019) Embedding training within warnings improves skills of identifying phishing webpages. *Hum Factors* 61(4):577–595
70. Flores WR, Holm H, Nohlberg M, Ekstedt M (2015) Investigating personal determinants of phishing and the effect of national culture. *Inf Comput Secur* 23(2):178–199
71. Dodge R, Coronges K, Rovira E (2012) Empirical benefits of training to phishing susceptibility. In: Gritzalis D, Furnell S, Theoharidou M (eds) *Information security and privacy research*, vol 376. Springer, Berlin, pp 457–464
72. Moody G, Galletta D, Walker J, Dunn B (2011) Which phish get caught? An exploratory study of individual susceptibility to phishing. In: *International conference on information systems 2011, ICIS 2011*, vol 3
73. Workman M (2008) Wisecrackers: a theory-grounded investigation of phishing and pretext social engineering threats to information security. *J Am Soc Inf Sci Technol* 59(4):662–674
74. Abbasi A, Zahedi FM, Chen Y (2016) Phishing susceptibility: the good, the bad, and the ugly. In: *2016 IEEE conference on intelligence and security informatics (ISI)*, pp 169–174
75. Kumaraguru P, Cranshaw J, Acquisti A, Cranor L, Hong J, Blair MA, Pham T (2009) School of phish: a real-world evaluation of anti-phishing training. In: *Proceedings of the 5th symposium on usable privacy and security. SOUPS '09*, ACM, New York, NY, USA, pp 3–1312
76. Orunsolu AA, Sodiya AS, Akinwale AT, Olajuwon BI, Alaran MA, Bamgboye OO, Afolabi OA (2017) An empirical evaluation of security tips in phishing prevention: a case study of Nigerian banks. *Int J Electron Inf Eng* 6(1):25–39
77. Steyn T, Kruger HA, Drevin L (2007) Identity theft—empirical evidence from a phishing exercise. In: Venter H, Eloff M, Labuschagne L, Eloff J, von Solms R (eds) *New approaches for security, privacy and trust in complex environments*. Springer, Boston, pp 193–203
78. Iuga C, Nurse JRC, Erola A (2016) Baiting the hook: factors impacting susceptibility to phishing attacks. *Hum-centric Comput Inf Sci* 6(1):8
79. Andrić J, Oreški D, Kišasondi T (2016) Analysis of phishing attacks against students. In: *2016 39th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, pp 1423–1429
80. Vishwanath A, Harrison B, Ng YJ (2016) Suspicion, cognition, and automaticity model of phishing susceptibility. *Commun Res* 45(8):1146–1166
81. Vishwanath A, Herath T, Chen R, Wang J, Rao HR (2011) Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decis Support Syst* 51(3):576–586
82. Kumaraguru P, Sheng S, Acquisti A, Cranor LF, Hong J (2008) Lessons from a real world evaluation of anti-phishing training. In: *2008 eCrime researchers summit*, pp 1–12
83. Alseadoon I, Chan T, Foo E, Nieto J (2012) Who is more susceptible to phishing emails? A Saudi Arabian study. In: *ACIS 2012: proceedings of the 23rd Australasian conference on information systems*

84. Harrison B, Svetieva E, Vishwanath A (2016) Individual processing of phishing emails: how attention and elaboration protect against phishing. *Online Inf Rev* 40(2):265–281
85. Vishwanath A (2015) Examining the distinct antecedents of e-mail habits and its influence on the outcomes of a phishing attack. *J Comput Mediat Commun* 20(5):570–584
86. Canfield CI, Fischhoff B, Davis A (2016) Quantifying phishing susceptibility for detection and behavior decisions. *Hum Factors* 58(8):1158–1172
87. Harrison B, Vishwanath A, Ng YJ, Rao R (2015) Examining the impact of presence on individual phishing victimization. In: 2015 48th Hawaii international conference on system sciences, pp 3483–3489
88. Parsons K, McCormac A, Pattinson M, Butavicius M, Jerram C (2013) Phishing for the truth: a scenario-based experiment of users' behavioural response to emails. In: Janczewski LJ, Wolfe HB, Shenoi S (eds) *Security and privacy protection in information processing systems*. Springer, Berlin, pp 366–378
89. Petelka J, Zou Y, Schaub F (2019) Put your warning where your link is: improving and evaluating email phishing warnings. In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. CHI '19. Association for computing machinery, New York, NY, USA
90. Tian CA, Jensen ML (2019) Effects of emotional appeals on phishing susceptibility. In: *Proceedings of the 14th Pre-ICIS workshop on information security and privacy*
91. Lee HS, Jeong DN, Lee SI, Lee SH, Kim KH, Lee HY, Cho HJ, Choi SW, Ko T (2019) Result and effectiveness of malicious e-mail response training in a hospital. *Stud Health Technol Inform*. <https://doi.org/10.3233/shiti90732>
92. Hermogenes MGG, Capariño ET (2019) Evaluating internet security awareness and practices of bulsu-sc students. In: *Proceedings of the 2019 7th international conference on information and education technology*. ICET 2019, Association for Computing Machinery, New York, NY, pp 62–66
93. Anawar S, Kunasegaran DL, Mas'ud MZ, Zakaria NA (2019) Analysis of phishing susceptibility in a workplace: a big-five personality perspectives. *J Eng Sci Technol* 14(5):2865–2882
94. Musuva P, Chepken C, Getao K (2019) A naturalistic methodology for assessing susceptibility to social engineering through phishing. *Afr J Inf Syst* 11:2
95. Jones HS, Towse JN, Race N, Harrison T (2019) Email fraud: the search for psychological predictors of susceptibility. *PLoS ONE* 14(1):0209684–0209684. <https://doi.org/10.1371/journal.pone.0209684>
96. Williams EJ, Polage D (2019) How persuasive is phishing email? the role of authentic design, influence and current events in email judgements. *Behav Inf Technol* 38(2):184–197
97. Shakela V, Jazri H (2019) Assessment of spear phishing user experience and awareness: an evaluation framework model of spear phishing exposure level (spel) in the namibian financial industry. In: 2019 international conference on advances in big data, computing and data communication systems (icABCD), pp 1–5
98. Lin T, Capecci DE, Ellis DM, Rocha HA, Dommaraju S, Oliveira DS, Ebner NC (2019) Susceptibility to spear-phishing emails: effects of internet user demographics and email content. *ACM Trans Comput Hum Interact* 26(5):1–28
99. Zielinska OA, Tembe R, Hong KW, Ge X, Murphy-Hill E, Mayhorn CB (2014) One phish, two phish, how to avoid the internet phish: analysis of training strategies to detect phishing emails. *Proc Hum Factors Ergon Soc Annu Meet* 58(1):1466–1470
100. Flores WR, Holm H, Svensson G, Ericsson G (2014) Using phishing experiments and scenario-based surveys to understand security behaviours in practice. *Inf Manage Comput Secur* 22(4):393–406
101. Greene K, Steves M, Theofanos M, Kostick J (2018) User context: an explanatory variable in phishing susceptibility. In: *Proceedings 2018 workshop on usable security*. Internet Society, San Diego, CA
102. Yang W, Xiong A, Chen J, Proctor RW, Li N (2017) Use of phishing training to improve security warning compliance: evidence from a field experiment. In: *Proceedings of the hot topics in science of security: symposium and bootcamp*. HoTSoS. ACM, New York, pp 52–61
103. Egelman S, Cranor L, Hong J (2008) You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In: *SIGCHI conference on human factors in computing systems*, pp 1065–1074
104. Neupane A, Rahman ML, Saxena N, Hirshfield L (2015) A multi-modal neuro-physiological study of phishing detection and malware warnings. In: *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security—CCS '15*. ACM Press, Denver, Colorado, pp 479–491
105. Hong KW, Kelley CM, Tembe R, Murphy-Hill E, Mayhorn CB (2013) Keeping up with the joneses: assessing phishing susceptibility in an email task. *Proc Hum Factors Ergon Soc Annu Meet* 57(1):1012–1016
106. Welk AK, Hong KW, Zielinska OA, Tembe R, Murphy-Hill E, Mayhorn CB (2015) Will the "phisher-men" reel you in?: assessing individual differences in a phishing detection task. *Int J Cyber Behav Psychol Learn* 5(4):1–17
107. Halevi T, Memon N, Nov O (2015) Spear-phishing in the wild: a real-world study of personality, phishing self-efficacy and vulnerability to spear-phishing attacks. *SSRN Electron J*. Accessed 29 Apr 2019
108. Kirlappos I, Sasse MA (2012) Security education against phishing: a modest proposal for a major rethink. *IEEE Secur Priv* 10(2):24–32
109. Kumaraguru P, Rhee Y, Acquisti A, Cranor LF, Hong J, Nunge E (2007) Protecting people from phishing: the design and evaluation of an embedded training email system. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp 905–914
110. Jackson C, Simon D, Tan D, Barth A (2017) An evaluation of extended validation and picture-in-picture phishing attacks. Microsoft Research (2007). Accessed 19 Dec 2017
111. Dhamija R, Tygar JD, Hearst M (2006) Why phishing works. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. CHI '06, ACM, New York, pp 581–590
112. Arachchilage NAG User-centred security education: a game design to thwart phishing attacks. [arXiv:1511.03459](https://arxiv.org/abs/1511.03459) [cs]. Accessed 29 Apr 2019
113. Lemay DJ, Basnet RB, Doleck T (2020) Examining the relationship between threat and coping appraisal in phishing detection among college students. *J Internet Serv Inf Secur*. 10(1):38–49
114. Bin Othman Mustafa MS, Kabir MN, Ernawan F, Jing W (2019) An enhanced model for increasing awareness of vocational students against phishing attacks. In: 2019 IEEE international conference on automatic control and intelligent systems (I2CACIS), pp 10–14

115. Li Y, Xiong K, Li X (2019) Understanding user behaviors when phishing attacks occur. In: 2019 IEEE international conference on intelligence and security informatics (ISI), p 222
116. Baral G, Arachchilage NAG (2019) Building confidence not to be phished through a gamified approach: conceptualising user's self-efficacy in phishing threat avoidance behaviour. In: 2019 cybersecurity and cyberforensics conference (CCC), pp 102–110
117. Wen ZA, Lin Z, Chen R, Andersen E (2019) What.hack: engaging anti-phishing training through a role-playing phishing simulation game. In: Proceedings of the 2019 CHI conference on human factors in computing systems. CHI '19. Association for Computing Machinery, New York
118. Yu K, Taib R, Butavicius MA, Parsons K, Chen F (2019) Mouse behavior as an index of phishing awareness. In: Lamas D, Loizides F, Nacke L, Petrie H, Winckler M, Zaphiris P (eds) Human–computer interaction—INTERACT 2019. Springer, Cham, pp 539–548
119. Pfeffel K, Ulsamer P, Müller NH (2019) Where the user does look when reading phishing mails—an eye-tracking study. In: Zaphiris P, Ioannou A (eds) Learning and collaboration technologies designing learning experiences. Springer, Cham, pp 277–287
120. Patel P, Sarno DM, Lewis JE, Shoss M, Neider MB, Bohil CJ (2019) Perceptual representation of spam and phishing emails. *Appl Cogn Psychol* 33(6):1296–1304
121. Akhawe D, Felt AP (2013) Alice in Warningland: a large-scale field study of browser security warning effectiveness. In: USENIX security symposium, vol 13
122. Khonji M, Iraqi Y, Jones A (2013) Phishing detection: a literature survey. *IEEE Commun Surv Tutor* 15(4):2091–2121
123. Purkait S (2012) Phishing counter measures and their effectiveness—literature review. *Inf Manage Comput Secur* 20(5):382–420
124. Afroz S, Greenstadt R (2011) PhishZoo: detecting phishing websites by looking at them. In: 2011 IEEE fifth international conference on semantic computing, pp 368–375
125. Pattinson M, Jerram C, Parsons K, McCormac A, Butavicius M (2012) Why do some people manage phishing e-mails better than others? *Inf Manage Comput Secur* 20(1):18–28
126. Karakasilotis A, Furnell SM, Papadaki M (2006) Assessing end-user awareness of social engineering and phishing. In: Proceedings of 7th Australian information warfare and security conference. Accessed 29 Apr 2019
127. Butavicius M, Parsons K, Pattinson M, McCormac A (2016) Breaching the human firewall: social engineering in phishing and spear-phishing emails. *arXiv e-prints*, 1606–00887. [arXiv:1606.00887](https://arxiv.org/abs/1606.00887)
128. Canova G, Volkamer M, Bergmann C, Borza R (2014) NoPhish: an anti-phishing education app. In: Security and trust management. Lecture notes in computer science, Springer, Cham, pp 188–192
129. Jansen J, Leukfeldt R (2015) How people help fraudsters steal their money: an analysis of 600 online banking fraud cases. In: 2015 workshop on socio-technical aspects in security and trust, pp 24–31
130. Parsons K, Butavicius M, Pattinson M, Calic D, McCormac A, Jerram C (2015) Do users focus on the correct cues to differentiate between phishing and genuine emails? Australasian Conference on Information Systems. [arXiv:1605.04717](https://arxiv.org/abs/1605.04717)
131. Turner CMB, Turner CF (2019) Analyzing the impact of experiential pedagogy in teaching socio-cybersecurity: cybersecurity across the curriculum. *J Comput Sci Coll* 34(5):12–22
132. Takata T, Ogura K (2019) Confront phishing attacks—from a perspective of security education. In: 2019 IEEE 10th international conference on awareness science and technology (ICAST), pp 1–4
133. House D, Raja MK (2019) Phishing: message appraisal and the exploration of fear and self-confidence. *Behav Inf Technol*. <https://doi.org/10.1080/0144929X.2019.1657180>
134. Eaton V, Cordova J, Greer T, Smith L (2019) A comparison of perceptions of cs majors and non-cs majors regarding email security. *J Comput Sci Coll* 34(3):31–37
135. Somestad T, Karlzén H (2019) A meta-analysis of field experiments on phishing susceptibility. In: 2019 APWG symposium on electronic crime research (eCrime), pp 1–14
136. Parsons K, McCormac A, Pattinson M, Butavicius M, Jerram C (2015) The design of phishing studies: challenges for researchers. *Comput Secur* 52:194–206
137. Wright R, Chakraborty S, Basoglu A, Marett K (2010) Where did they go right? understanding the deception in phishing communications. *Group Decis Negot* 19(4):391–416
138. Williams EJ, Hinds J, Joinson AN (2018) Exploring susceptibility to phishing in the workplace. *Int J Hum Comput Stud* 120:1–13
139. John OP, Srivastava S (1999) The big five trait taxonomy: history, measurement, and theoretical perspectives. In: Pervin LA, John OP (eds) Handbook of personality: theory and research, 2nd edn. Guilford Press, New York, pp 102–138
140. Rawlinson G (2007) The significance of letter position in word recognition. *IEEE Aerosp Electron Syst Mag* 22(1):26–27
141. Stanton B, Theofanos MF, Prettyman SS, Furman S (2016) Security fatigue. *IT Prof* 18(5):26–32
142. Crossler RE, Johnston AC, Lowry PB, Hu Q, Warkentin M, Baskerville R (2013) Future directions for behavioral information security research. *Comput Secur* 32:90–101
143. Rajivan P, Gonzalez C (2018) Creative persuasion: a study on adversarial behaviors and strategies in phishing attacks. *Front Psychol* 9:135
144. Nguyen TH, Yang R, Azaria A, Kraus S, Tambe M (2013) Analyzing the effectiveness of adversary modeling in security games. In: Proceedings of the twenty-seventh AAAI conference on artificial intelligence. AAAI'13, AAAI Press, Bellevue, Washington, pp 718–724
145. Polley S (2017) ComBAT phishing with email automation. SANS Institute Information Security Reading Room 29. <https://www.sans.org/reading-room/whitepapers/email/combat-phishing-email-automation-38025>
146. Dwork C (2008) Differential privacy: a survey of results. In: Agrawal M, Du D, Duan Z, Li A (eds) Theory and applications of models of computation. Lecture notes in computer science. Springer, Berlin, pp 1–19

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.