REPHRAIN

Protecting citizens online

"I didn't click": What users say when reporting phishing

Nikolas Pilavakis, University of Edinburgh

Adam Jenkins, University of Edinburgh

Nadin Kokciyan, Univeristy of Edinburgh

Kami Vaniea, University of Edinburgh

March 2023

UK Research and Innovation · University of Bristol · UCL · The University of Edinburgh · King's College London · University of Bath

"I didn't click": What users say when reporting phishing

Nikolas Pilavakis
University of Edinburgh
nikolaspilavakis@hotmail.com

Adam Jenkins
University of Edinburgh
adam.jenkins@ed.ac.uk

Nadin Kökciyan
University of Edinburgh
nadin.kokciyan@ed.ac.uk

Kami Vaniea
University of Edinburgh
kvaniea@inf.ed.ac.uk

***Abstract*—When people identify potential malicious phishing emails one option they have is to contact a help desk to report it and receive guidance. While there is a great deal of effort put into helping people identify such emails and to encourage users to report them, there is relatively little understanding of what people say or ask when contacting a help desk about such emails. In this work, we qualitatively analyze a random sample of 270 help desk phishing tickets collected across nine months. We find that when reporting or asking about phishing emails, users often discuss evidence they have observed or gathered, potential impacts they have identified, actions they have or have not taken, and questions they have. Some users also provide clear arguments both about why the email really is phishing and why the organization needs to take action about it.**

I. Introduction

Phishing incidents occur when an attacker attempts to trick someone via email into providing sensitive data, such as account information, passwords, and banking details. For organizations, phishing is a common gateway attack as it can be used to gain initial access to protected internal systems which can then be combined with other exploits to further access sensitive data and critical systems [64], [65]. This approach is both very effective and very expensive. Phishing losses amounted to £53.7 million due to impersonation fraud [21] in the United Kingdom in just 2020. Due to these risks, organizations take phishing seriously and engage in protective behaviors such as deploying automated scanning of emails and training their staff on how to identify and report phishing [51]. In 2017, organizations were spending an average of $290 thousand every year on training [59].

Reporting phishing incidents is one of the best ways to enable organizations to protect their staff from an attack by removing it from inboxes or blocking linked websites. However, we know surprisingly little about reporting since there is limited research on the effective use of phishing reporting processes. Various approaches focus on managing phishing in organizations such as automation [20] and staff training [35], [37]. While it is common for organizations to lament their staff's tendency to interact with phishing, the truth is that humans have a wide range of abilities, including attentiveness [18], [12]. Given the likelihood that a large scale phishing campaign will result in some people interacting with it, it is equally likely that other people will identify it. If those people report the phishing quickly, then the organization can take preventive measures like removing it from all staff email inboxes and blocking malicious links using a firewall [4]. In other words, it only takes one person reporting a phishing email to protect a whole organization from it. Consequently, most organizations have a process in place where users can report phishing for Information Technology (IT) staff to act on [41], and having a reporting process in place is considered best practice in government advice [42], [26]. These practices often focus on the technical aspects and their responses, such as removing identified emails from inboxes, with far less focus given to users who are reporting. Organizations may have some policies around sending a response to those who report, or providing some standard guidance, but currently we have limited knowledge about what kinds of information and questions people provide alongside their reports.

We believe that organizations would strongly benefit from a deeper understanding of how to support their users in identifying phishing as well as how to encourage them to report phishing. Existing research has explored cues and approaches users take to identify phishing through a range of methods including interviews [19], think aloud lab studies [30], eye tracking [3], and surveys [68] to understand what they look for when examining potential phishing messages. These works produced interesting results but most of them focus on either recalling past experiences or studies with lab-curated data sets and tasks. There are few works which explore what people notice when identifying phishing messages in the wild (e.g. [9]) and most of these are still conducted against mock phishing emails. Similarly, there is minimal research on the support required by people who may have identified potential phish but are still trying to determine if it is legitimate or not.

In our own prior work with a University [4], we noticed that help desk staff were often overwhelmed with reports, making it challenging to read and respond to each one individually. Instead staff make use of the fact that reports are primarily created by forwarding the phishing email and therefore have the email's subject line, which is often similar throughout a single campaign. Staff use this observation to bulk respond to all similar-looking phishing reports at once, often by using pre-written standard wording which includes stating if the email is phishing, asking the user to not click any links, requesting that they delete the email, and instructions on how to reset their password if they had engaged with the email. While this approach is likely effective in most cases, it made us

Symposium on Usable Security and Privacy (USEC) 2023
27 February 2023, San Diego, CA, USA
ISBN 1-891562-91-6
https://dx.doi.org/10.14722/usec.2023.233129
www.ndss-symposium.org, https://www.usablesecurity.net/USEC/

wonder what information and queries were being ignored in the process. It also made us consider if the standard guidance provided aligned well with what users were talking and asking about in their reports.

In this paper, we consider communications users have with a University help desk about potential phishing incidents. Our goal is to use such reports to better understand how people are thinking about phishing in the wild and the types of support they request to help them manage phishing-related issues. We ask the following three research questions in regards to contacting a University help desk about phishing:

- **RQ1** What statements and information do people provide when submitting phishing reports?
- **RQ2** What questions do people ask and what kinds of support requests do they make when submitting phishing reports?
- **RQ3** How well do the topics mentioned in questions align with the information provided in the standard response?

To answer these questions, we qualitatively coded a randomly selected set of 270 phishing-related University help desk tickets which spanned a period of nine months. Additionally, we provide some general descriptive statistics, including an automated analysis of report subject lines, on a larger set of all 984 reports made during the same period to better understand who is reporting and the types of reports being made. Finally, we discuss how the standard advice given to users who report differs and aligns with what users are asking about.

We find that 57.0% of the analyzed phishing reports contained information beyond a simple report or query about if the reported email really was phishing. Users provided *evidence* of their reasoning such as citing that the *From* address was inconsistent with expectations; they highlighted the *impact potential* of an email by pointing out aspects like the number of people impacted or if it impersonated an authority figure such as the Head of School; they listed *actions* they had and had not taken such as stating that they had not clicked any links; and finally they asked *questions* to clarify what steps they should take next or better understand the problem they were having such as asking if changing all their passwords was truly necessary. The results highlight the range of phishing aspects users are taking into account as well as problems they are facing that are currently poorly covered by existing phishing guidance. Based on our results, we conclude with four main areas to tackle in the future to improve phishing report processes in the organizations.

II. Related Work

The topic of phishing is well studied in the literature with several systematization of knowledge papers already written on the topic [22], [6]. To highlight the gaps that our research and findings fill, we first detail the numerous studies which highlight the features, both technical and contextual factors, that end-users and experts use to identify phishing. We then contextualise our research by detailing the range of defences that are used by organizations to defend against phishing, such as the technical filters used or the mechanisms focused on end-users. This includes systems designed to help highlight phishing and the training schemes which are common place within industry. Our work focuses on a rarely investigated aspect of the organizational phishing defence ecosystem, the reporting component.

A. Users Identifying Phish

Identifying phishing emails is an immensely difficult task as they are specifically designed to mislead, with a number of studies investigating strategies used by both phishers and users. Downs et al. [19] interviewed 20 non-expert users to identify the strategies they used when encountering potential phish. Several cues were reported including spoofed "from" addresses and unexpected or strange URLs. The strategies employed by participants required inspection of email contents for good grammar and the perceived relevancy (e.g. is this email for me?). Dhamaja et al. [18] examined users ability recognising phishing websites, identifying a number of deceptive techniques such as substituting similar letters into the domain name or mimicking the aesthetic of a spoofed organization. Participants relied on strategies, such as reviewing a pages content and domain name, which were found to be unsuccessful with phish sites able to fool more than 90% of participants. Jakobsson et al. [30] used a think-aloud protocol to identify features of phish which made them appear more authentic, they found that emails that contain personalization, including data readily available to the public, were perceived as more trustworthy by users.

An abundance of research has investigated different factors and how they correlate with users' susceptibility to falling for phishing [17]. Analysis of demographics factors including age, find that older users are more likely to fall for phish than their younger counterparts [46], [39], similarly gender, with female users more likely to fall for phish than males [46], [39], [10], [29]. Sheng et al. [60] used the Mechanical Turk (mTurk) platform to study the relationship between susceptibility and these demographic factors by surveying 1001 participants. Through use of mediation analysis, the authors state that observed gender differences may be due to differences in technical ability within these compared demographics. In this paper, we are concerned about what the users say when they report phish within a singular organizational context, that of a University, which are known to have a wide range of demographics but skews towards those with higher educational qualifications.

Many spoofing techniques can be employed by malicious actors to trick or manipulate users into interacting with phishing, with techniques continuing to evolve. This makes the identification process for end-users a continually moving target as attackers adapt to systems and abilities of their targeted users and organizations. Blythe et al. [11] provided a multi-method collection of four studies and found that the inclusion of spoofed logos were difficult for participants to recognise. Interestingly, the authors found that blind users had developed robust strategies focusing on email content, indicating that the visual features may not be as necessary as previously thought. Work by Alsharnouby et al. [3] complimented previous studies by using eye tracking to collect data on the visual cues participants paid attention to when evaluating suspicious phishing websites. Their results showed that despite priming, participants were only able to identify 53% of legitimate phishing on average. Participants spent little time viewing

security indicators when assessing websites, instead using similar strategies to those found in prior work [19], [18], such as close inspection of web-page content or paying attention to the URL, despite misunderstandings regarding syntax such as differences between domains and sub-domains. These results hold when we also consider that many end-users struggle to understand the syntax and function of URLs [2], [5], [53]. More recently, Zheng and Becker [70] conducted a study to investigate the impact of presenting email headers to end-users had on their ability to identify legitimate phishing examples. Their results are rather surprising with users unable to use such header features to correctly identify phishing emails, therefore implying that highlighting such features to end-users would not improve the overall state of phishing detection. Similarly, our analysis on phish reports shows that users do not mention header features when they provide some evidence about why they think the message could be phish.

Research which has ventured beyond the confines of a lab or synthetic phish are rare, however Benenson et al. [9] used a field experiment involving 1200 university students and simulated phishing messages, which were sent through either email or Facebook. Participants were more likely to click on the link if on Facebook compared to email. Additionally, they found that a number of reasons were given for clicking on link, such as curiosity or assumptions regarding known sender. Conversely, participants would avoid the link if suspicious of context the message was received or respect for others privacy as the message appeared to not be for themselves. Our work explicitly extends knowledge by analysing a *real world data set* of phishing reports made by users to the help desk of a University, giving great validity to the observations we present. Moreover, analyzing such reports helps us in understanding the users who identify and report phish in organizations.

Phishing is not relegated to an issue for end-users, as research has shown that even so-called experts in technology and cyber security can fall for the tricks that they employ. Wash [67] interviewed 21 IT experts to identify the process they go through to recognise phishing emails. He identifies three stages used, beginning with the phish being treated like any other email, until they notice inconsistencies, further alarming the participants. Once sufficient evidence has been spotted, they will become suspicious and will actively search for signs of phish, such as hovering over any links or opening email headers. By doing so they will come to a decision regarding its legitimacy and act, with the majority deleting the email. Wash et al. [68] extended this research by surveying 297 users for descriptions of their interactions with phish to identify their approaches to phishing detection. Similarities existed between non-experts and experts, with the authors arguing that users bring unique knowledge regarding their own expected email and situation.

B. Technical Protections

Technical barriers have been developed to prevent successful phishing attacks, such as preventative measures which filter phish from reaching users' inboxes [20]. These technical implementations can take a number of forms including simple heuristics [16], such as features of the URLs found within the email [43]. Other passive filtering approaches incorporate allow lists [13] and deny lists [62] which compare potential phish against known lists of legitimate phishing examples. In recent years, we have also seen the use of more sophisticated approaches such as machine learning and deep learning techniques to identify phishing attacks [50], to prevent business email compromise attacks [15] and to detect URL-based phishing attacks [55]. Stringhini and Thonnard focus on modelling the email-sending behaviour of the users in order to fight spear-phishing threats [63]. Although these methods have been shown to be highly effective, they unlikely to ever reach 100% accuracy due to fact that phish are intentionally designed to bypass these technical barriers and deceive end-users. Hence, human judgement is considered an essential part of an organization's overall security posture [23].

C. Phishing Detection Support

Due to the essential role that users play in the detection of phish, a range of interventions have been designed. Franz et al. [22] provided a systematization of user-centred phishing interventions, identifying four categories: education, training, awareness-raising and design. Education and training aim to improve users knowledge and skills and promote long term security behaviours. The latter pair cover methods that guide users through a specific context.

1) Design & Awareness-Raising: Awareness-raising interventions are usually constructed for a specific context to highlight potential threats, such and highlighting features of URLs in messages found in users' emails [66]. Approaches in highlighting URLs has been found to be more effective than the use of email warning banners [47]. Design interventions are more passive and relate to the design decisions made, which instigate desired secure behaviours, such as highlighting sender's information [44], or browser add-ons which provide warnings regarding web pages visited [54].

2) Training & Education: Education is considered a vital component of protection against harms and has long been an area of interest from the academic community [27], [36]. A range of options for user training are available including game-based training [12], [8], [61], real-time solutions [7], [69], and the use of simulated phishing attacks embedded into users' working contexts [36], [37]. However, the execution of these phishing drills requires extensive efforts by system administrators in crafting examples and handling user reports [4], [33]. Additionally, despite the repeated advice of reporting spotted phish [26], [42] users may still not do so. Kwak et al. [38], investigated why this was the case and identified that many users lacked self-efficacy regarding their ability to assess the legitimacy of phish. In other words, they were not confident enough to report the phishing in case they might be wrong. Interestingly, the authors also noted that participants stated that they were unlikely to report phishing emails then believed to be obvious examples, which is better than nothing but prevents examples being used to improve technical defences.

The work of Reinheimer et al. [52] evaluated the effectiveness of phishing training and education, finding that participants had a significant improvements to their ability to identify examples of phishing emails immediately after and up to four months after training had taken place. However, this improvement had dissipated by around six months indicating that repeated updates are necessary to see continual benefits of phishing training.

III. Methodology

This paper is part of a larger project to understand how technical staff and end users at the studied University engage with phishing emails, particularly in regards to reporting them. Earlier work included a set of in-depth interviews with technical and support staff aimed at understanding the phishing protections used as well as how reported phishing is handled [4]. This paper is an extension of that earlier work. It was inspired by comments from technical staff that users are sometimes frustrated when their report includes a question, but that question is not answered. Technical staff normally bulk respond to all reports with a similar subject line (a.k.a. a campaign) with a single standard response, because reading every report would take a great deal of time. Hence, the aim of this paper is to better understand what users are saying when they report phishing and what kind of support they expect. The methodology throughout has been informed by the existing collaboration with technical staff to ensure both respect for users and accuracy in terms locating and interpreting phishing reports.

A. University Processes

While different from industry organizations, Universities have some unique features that make them interesting to study in regards to phishing. Unlike a company where all employees can be required to complete training, Universities are populated by a wide range of people including temporary workers, students, professional services staff and academic staff who are hard to compel to take training. The turnover for a University is also quite large with many students joining and leaving every year. Many staff at Universities also communicate regularly with people external to the organization including doing things like clicking links and opening attachments from unknown senders which makes automated protections harder to implement. These issues are possibly why the sector of Education has the highest click-through rates [64].

The University user guidance on reporting phishing is to contact the main help desk which uses a ticket tracking system. The guidance lacks any information on what responses the reporter is to expect, nor does it state explicitly what the phishing email will be used for. Nearly all phishing reports are recorded as tickets including the initial request, the reported email, discussions between staff, and any response that was sent back. In reality, the University has specialized help desks in some departments and users will sometimes report there instead of to the main help desk; however, we observed that staff in these specialized help desks do forward on phishing reports they consider to be problematic on to the main help desk system. So the main help desk ticket system is the main method for users to report phishing at the University.

To understand what users are saying when they report phishing, we consider 984 tickets collected over approximately nine months and then further qualitatively analyze 270 tickets to gain a more in-depth understanding of what people say and ask when reporting phishing. The research was approved for ethics through our School's research ethics board, and we worked with the technical staff of the studied University throughout to ensure they were comfortable with our research. Additionally, due to our close relationship with the IT staff and services we have continually used our research findings to feedback to their phishing reporting process, with the aim of improving the quality of the service for both reporters and staff who handle the reports. Measures were taken to protect the identity of those who reported, with quotes throughout the paper being carefully selected to ensure that they respect users' privacy.

B. Dataset

While the ticket tracking system records all communications with the help desk, phishing tickets are not given any specific tag or label in the system, making them hard to accurately locate. To find the phishing tickets we used two approaches. First, we looked for any tickets that had "phishing", "scam", or "spam" in the request as these terms are often used by reporters, and the help desk sometimes adds "phishing" to ticket subject lines that are unclear. Secondly, the help desk also has a set of "standard solutions" which are pre-written responses they provide for common requests, including phishing, so we also searched for any tickets containing the word "phishing" in the response as this term appears in all standard solutions. Our contact at the help desk confirmed that these search terms are consistent with what they observe and their procedures.

We limited our search to tickets created between the 27th of October 2020 and 2nd of August 2021, resulting in 984 tickets. In mid-October 2020 the University introduced an automatically added banner to the top of all incoming emails originating from non-University domains. The banner states that the email is not from the University and that the recipient should be careful when clicking. Many UK universities introduced a similar banner around this time likely due to Office 365 providing an easy-to-use feature to do so. This feature was rolled out across the University over a couple of weeks and our contact confirmed that the roll-out had completed by the 27th of October. The data collection time frame was selected to start slightly after the banner was introduced and end just before the start of the next school year.

1) Reports of non-phish: We were also interested in cases where a user reported phishing but the response from the help desk indicated that the reported email was not phishing or otherwise malicious. To identify these cases we first identified commonly occurring words appearing in standard solutions such as "genuine", "legitimate", "safe", or "not * phishing" where * matched any white space or word. We then used a set of regular expressions to identify the help desk responses to phishing tickets matching these expressions. This approach returned 94 tickets out of the 984 tickets, which were then manually reviewed, we identified 22 emails which the help desk identified as not phishing.

2) People who report: In the full set of 984 reports there were a total of 633 unique reporters of which 497 had only filed a single report and 86 had filed two. A Professor in the social sciences had provided 71 reports with the next most frequent reporter being a member of the University's computer security team who provided 14 reports. Only 82 reports were made by members of the University IT services group. The University is also made up of a number of colleges all of which are well represented in the data, though amusingly the

college associated with computer science had the lowest levels of reporting among the colleges. In other words, most people reported only one or two times, and the people who report are spread widely through the University with only a few people being frequent reporters.

C. Subject Line Analysis of Reports

When a ticket is submitted via email, the subject line of the email becomes the subject line of the ticket which is then the most visible piece of information to the first line help desk staff who process the reports. Our prior work found that help desk staff make heavy use of subject lines when processing phishing reports [4]. We therefore used automatic clustering of subject lines of the 984 reports to understand what type of subject lines appear in reports.

We started by preprocessing the subject lines to normalize the text being used in each case by removing stop-words and punctuation. We applied term frequency-inverse document frequency (TF-IDF) [48], [49] to get important words from the reports, where the value of a word is computed based on the word frequency in the subject line of a report and the popularity of the word in all the subjects in the corpus. We then grouped the reports based on subject similarity by using K-Means [40] clustering algorithm. The Elbow method [32] suggested grouping scenarios into five clusters. The five clusters identified are C_0(98), C_1(305), C_2(129), C_3(365) and C_4(87); where the number of instances belonging to that particular cluster is shown in parentheses.

In Table I, we report the identified five clusters by representing them using the top 10 most important terms (i.e., terms with the highest TF-IDF scores). C_0 mostly represents reports that have subject lines starting with 'Fwd:' (i.e., forwarded emails), or the reports that have been submitted via a form. The University help desk gives the users an option to submit requests via a form which asks them to choose from a predefined list of subjects (e.g., Email / Office 365), or the users can also choose the 'Other' option to provide a custom subject line for their report. Forwarded emails have the advantage of defaulting to the title of the original email, which if not edited can provide a brief description about the content of the email. In C_1, most of the subject lines include the words 'phishing', 'phishing attempt', 'suspicious email', 'internal looking phishing' or 'phishing email attached'; which are non-descriptive subjects about the content of the reported emails. In C_2, most of the subject lines include just the word 'spam', the remaining subject lines are accompanied by one or two more words (e.g., spam reporting, spam email). Similar to C_0, C_3 is also a cluster for forwarded emails that have subject lines starting with 'Fw:' this time. Phishing campaigns are visible in this cluster since a set of reports are being submitted with exactly the same subject line (e.g., 'Fw: You have a new voicemail', 'Fw: Your parcel is on hold', 'FW: New Update From *X*'). The keywords 'new' and 'update' also appear in the subject lines of reports in this cluster. C_4 is the cluster where we observe reports with subject lines including the word 'scam', most reports include subject lines such as 'scam email' or 'potential scam'.

According to this analysis, we observe that half of the reports include generic subject lines such as 'spam' or 'scam', which fall short in describing the issue raised as part of the ticket; while the forwarded messages have the potential to be more self-explanatory for the help desk who are trying to handle tickets.

D. Qualitative Coding Approach

The lead researcher read through a random sample of about 100 tickets out of the set of 984 tickets taking notes and memos [57] as they went. For qualitative coding, a random set of 300 tickets was sampled from the full set of 984 tickets. 300 were selected based on the lead researcher's observation that many reports were relatively short, often with only one or two words, so coding roughly a third of the set was reasonable. After the coding described below, two researchers involved in the coding process reviewed their memos and discussed if coding more was likely to impact the result. They decided that the codebook well matched the data and that within-code concepts were repeating.

The lead researcher started by building an initial codebook based on their initial observations. Both coders then attempted to use the codebook on 30 reports, discussed differences, and revised the codebook followed by the same with another 32 reports resulting in a stable codebook and agreed codes for the initial 62 reports. One coder then went through and coded the remaining 238 reports. The second coder coded the first 27 reports and last 31 reports so inter-coder reliability could be computed to track drift.

During the coding process, the coders identified 30 tickets which could not be classified as phishing reports. For example, a user contacted the help desk to complain about the warning banner being added to all their emails. These messages were removed from further analysis resulting in 270 tickets which are described in the following sections.

We computed agreement between two researchers by using Krippendorff's alpha [25], [34] with Jaccard's distance metric [28]. Krippendorff's alpha is a measure of inter-rater reliability widely used in content analysis that supports partial agreement when raters use multiple values on the same data unit; hence, it was a suitable measure for our study. The initial Krippendorff's alpha value was computed as 0.69 based on the the initial set of 27. The final Krippendorff's alpha was computed as 0.77, which indicates an acceptable reliability between coders [34].

E. Limitations

The presented research is a case study of a single University over about nine months. The researched time period also coincided with the COVID-19 pandemic so for the majority of the studied time frame University staff and students were instructed to work from home if possible. Like all case studies, this research attempts to accurately represent the studied case but the results may not generalize well to other organizations or situations. For example, phishing is reported at the studied University by emailing the help desk, another organization might have a button on the email client, which would make reporting easier (likely resulting in different levels of real phishing being reported and different information being provided). In another organization, employees may only be contacted in the case

TABLE I. CLUSTERS TOGETHER WITH THEIR TOP 10 TERMS WITH THE HIGHEST TF-IDF SCORES.

C_0	fwd, form, helpline, email, 365, office, phishing, information, notice, technology
C_1	phishing, email, attempt, emails, suspected, suspicious, attached, possible, internal, automatic
C_2	spam, email, suspected, phishing, fw, reporting, possible, mail, report, potential
C_3	fw, new, account, update, voicemail, notice, email, mailbox, information, notification
C_4	scam, email, potential, phishing, please, fw, possible, emails, closed, automatic

of false positives; whereas all the true positives would be considered to improve the spam filters.

We are also studying the reports people made to a University help desk, but everyone does not report phishing to such services, which means we are looking at a self-selected group of people who do not perfectly match the University's demographics. It is likely that people who report have above average technical skills and that they are also more confident or more worried than an average user encountering a phishing email. That said, we did see reporting from a wide range of individuals including staff and students across all the University colleges and other organization groups suggesting that these are genuine reports from actual users rather than reports from a small number of people who are highly experienced or who do such identification as part of their job.

IV. RESULTS

In this section, we first start with an analysis of non-phish reports, which highlights the common errors people make when they report phishing. We then introduce our qualitative results based on our analysis of 270 reports. The reports include content from five main high level codes, where each report can be assigned to multiple codes: (i) *Just reporting*, which mostly includes forwarded phishing emails or simple "is this phishing" requests; (ii) *Evidence*, which includes reports providing evidence and the reporter's reasoning before contacting the help desk; (iii) *Impact potential*, which includes reports saying why the phishing might negatively impact others; (iv) *Actions taken*, consists of reports that emphasize specific actions (not) taken by the reporters; and (v) *Questions*, that consists of reports including questions to the help desk such as the next steps to follow. Each high level code (except *Just reporting*) also includes a set of subcodes to represent reported phishing emails in a fine-grained manner. The content of the reported phishing emails is then matched against this codebook. Table II shows these primary codes together with the secondary codes, which detail the types of information included by the user. We also include the number of reports where a specific type of information appeared. For example, 'From address' has been mentioned in 53 phishing reports; where the reporters made comments about the sender of the original email. Note that the quotes in this section have been minimally edited for spelling and capitalization to improve readability, identifying information has also been redacted where appropriate.

TABLE II. TYPES OF INFORMATION MENTIONED WHEN DISCUSSING PHISHING WITH THE HELP DESK. THE TABLE SHOWS THE COUNT OF THE NUMBER OF REPORTS THAT CONTAINED THE SPECIFIED TYPE OF INFORMATION, A REPORT CAN BE ASSIGNED TO MULTIPLE CODES.

Code	Subcode	Count
Just reporting		116
Evidence		82
	From address	53
	Cues	15
	Technical	12
	Unsolicited	11
	Other people	9
	Banner	6
	Tools	5
	Other evidence	2
Impact potential		62
	Repeated emails	21
	Compromised	18
	Convincing	15
	IT systems	8
	Number targeted	7
	Other impacts	0
Actions taken		85
	Not clicked	35
	Clicked	28
	Deleted	19
	Changed login	8
	Gave data	7
	Did not give data	7
	Did not open	7
	Responded	1
	Did not respond	2
	Opened	0
	Other action	2
Questions		33
	Next Steps	11
	Other Questions	22

A. Reports of non-phish

People who reported phishing were quite accurate. Only 22 out of the full 984 tickets had a response from the help desk informing the reporter that the email was not phishing. When we inquired, help desk staff also confirmed that inaccurate reports are rare. There are several possible causes, the most likely is that reporting to the help desk requires the reporter to first identify that the help desk is the best place to report to and locate their email address. While the required effort is not large, it may be enough to prevent casual reporting such that people only report if they do so often or if they feel strongly about the email they are reporting. It may also be that only people who have high confidence are willing to report, which may also account for the high accuracy.

Messages from reporters on non-phish reports were very similar to the comments on accurate ones. Some users were highly confident that these were indeed phishing while others had similar "is this phishing?" comments seen on the other reports. The larger difference was in the responses. Help desk staff put effort into verifying these reports, sometimes contacting the claimed sender to verify the email's validity or asking the security team to look into the case. Responses

to reporters were also highly personalized, none of the 22 reporters were sent straight standard solution text and all were customized to the content that had been sent. When the email was from a University entity, the advice often stated this clearly and conveyed that the email was safe to interact with. But for non-University emails, the advice was usually more hesitant and aimed at helping the user make a decision such as commenting that the email looked legit or recommending that if they were unsure they should call or check their accounts via the company's main website and not use links in the email. "Better safe than sorry" was a common sentiment when giving advice about non-University emails.

B. Just reporting (116, 43.0%)

By far the most common type of interaction was when the person simply forwarded a phishing message with little to no extra added commentary (as also observed in Section III-C in subject lines). Sometimes these were simple forwards with nothing added. But often they would have a single word like "phishing" or "spam" added to the front of the subject line or top of the email. More verbose reporters would add a simple message such as: "here is another one for you" or "this looks like phishing".

This code also included people who were asking if the email was phishing with no additional questions or information. While some of these queries were quite direct in asking for feedback (e.g. "Is this phishing?"), others were more implied (e.g. "I am assuming that this is a scam") or terse (e.g. "phishing?"). After discussion, we coded all such queries as "Just Reporting" because ultimately all messages in this code were in some way suggesting that the email might be phishing and anyone reporting phishing is probably open to being told they are wrong.

Because this code was only applied in cases where no other questions or information was provided, it is functionally an exclusive code and was never dual coded with any other.

C. Evidence (82, 30.4%)

When reporting, some people included *evidence* or information that the person had considered before contacting the help desk. The evidence was often presented as part of an explanation of their reasoning either leading to an assertion that this was likely phishing, or a query for guidance because they could not be certain.

By far the most common type of evidence was the *From address*. People often mentioned that they did not know the sender, or would not expect communication from them. Or, conversely, they would talk about how they did know the sender or the sender's purported institution which caused them to take the email more seriously. Some people even put effort into looking the sender up either in the University directory or on their purported institution's page (e.g. "sender looks like a Chemistry student").

People also talked about *cues* that they picked up on that did not make sense or were out of character, often in terms of actions they had or had not previously taken that would have resulted in the email. Terms like "out of context" were used to describe the problem or "not expecting to get an email from". For example, someone reported a phishing message pretending to be an email bounce notification, and they commented how they had never sent the original email so the bounce message could not be legitimate. Similarly the following notes a discrepancy with timing: "the time the email was sent compared to the time the call was meant to be received do not align". Cues involving issues like spelling errors were only mentioned by one reporter.

A couple of people also reported phishing because the email looked problematic, and they were hesitant about the nature of the email as it misaligned with their current context. One person was expecting a package and got what looked like a package payment scam but was unsure. Similarly, another was "expecting a payment from HSBC" but "thought I would check first in case more users have received this?"

The other types of evidence were less common, though still interesting, and included: mentioning that *other people* they knew had also received the email indicating that it was illegitimate; talking about a *tool* that had produced an alarm, such as the browser phishing warning appearing; mentioning *technical* aspects of the email such as what server the email had originated from; and talking about the email warning *banner* the University adds to all external email.

D. Impact potential (62, 23.0%)

Reporters were also concerned about the impact the phishing emails could have on the University and others. They were concerned that the email might deceive other people and commented on aspects of it that might cause negative impacts to the wider University. Commentary in this code was often toned around helping others by getting the phishing blocked or by getting help for someone who had already been compromised. Reporters seemed aware that sending a message to the help desk could result in getting the email blocked for themselves and others.

1) Repeated Emails: The users also raised the issue of receiving repeated phishing emails in two different contexts. First, cases where the reporter had received multiple copies of the same phishing at roughly the same time, which had motivated them to report it. For example: "I have just received an email on both my student and personal account". Second, cases where the reporter had gotten similar emails across multiple days which had motivated them to report after seeing that the phishing was not being resolved. "I am forwarding one of the many identical emails that I have been receiving since Sunday." Some reports even mentioned repeats happening over long time periods: "I keep on getting the attached emails a few times a week."

An interesting point about reports involving repeated emails is that in most cases the person received multiple phishing emails before reporting any of them. Consider the following example:

> Received 4 of these this morning deleted 3 and thought I would sent you this one as this does look like a scam.

This observation suggests that people are not reporting all phishing they see to the help desk and instead only report

those that are particularly bothersome, impactful, or perceived as abnormal in some other way.

2) Compromised Accounts: This subcode often co-occurred with evidence comments about the *From address*. People would state or imply that the email was coming from a University email address that may have been compromised. Sometimes they would explicitly call out the account as compromised:

> I wanted to let you know about a possibly compromised account: that of [name and email address].

Or ask the help desk to assist the other user: "it looks like it's been sent from a proper [University] account, so can someone flag this with the user, as they may need to change passwords etc." Others were more terse but still called out the internal nature of the from address: "internal phishing email" or assumed that the attacker was just pretending to use an internal email: "This looks like a fishing email pretending to come from a uni address." A few also self-identified that they might be compromised and contacted asking for help: "The trail of failed email deliveries suggest my account has been compromised - please advise how to change password".

3) Convincing Emails: People also mentioned how convincing the email was as a way of emphasizing that it needed to be addressed quickly, often to protect others. The most common reference was to authority figures being impersonated such as heads of school, departments or other organizational units. The following quote is more eloquent than average, but touches on many of the concepts expressed by reports associated with this code.

> Our [Head of School] has been impersonated again by this address - [attacker email]
> In the past you have applied a rule to silently block this email on the mail server. Please could you do that again? Some people from our School have engaged and I'd like to cease comms from this address ASAP.

4) IT Systems: Emails purporting to be from the University's IT systems were also considered as potentially harmful and were pointed out with messages like: "looks like it is coming from you". This code is similar to the *convincing* code but instead tries to mimic University services (i.e. Office 365) or support teams such as the security team, email team, or the help desk. Some of the reports also include questions about if the email really was from the University or queries asking for reassurance that the threatened action would indeed not happen. One example is as follows:

> I received the below message from 'Microsoft' on Tuesday but I'm not sure whether to believe it or not. If its real I probably do need to look through the messages. However I don't want to click anything until I know for sure what is happening.

5) Number targeted: The number of people getting the email is also mentioned, usually in reference to a department, email list, or other people mentioning getting it. The focus here is normally on the scale of impact. Interestingly, some of the reports in this code were actually forwards of department-wide announcements about specific phishing campaigns. Evidently an attacker targeted a specific organizational group, resulting in an enterprising member of the group sending out a wide announcement warning people about the phishing email, this email was then forwarded on to the help desk to make sure they were aware of it. Again, this result indicates that users may not be reporting phishing and instead opting to resolve it locally by sending out local emails to warn others.

E. Actions taken (85, 31.5%)

People also reported the various actions they had or had not taken in relation to the phishing message. It is worth noting that the banner added to incoming external email advises people not to click links or open attachments if they are uncertain, and most of the standard solution text used by the help desk advises people who have interacted with phishing to consider changing their password and to run a virus scan on their computer. It is interesting to see that some of these concepts, such as *clicking*, appear frequently in reports while others like *virus scans* appear rarely.

When reporting phishing, users would sometimes add in comments about their own actions. These comments tended to either list the actions they had not taken (i.e. not clicked), or list the things that they had done which they were now worried about (i.e. entered login details).

When discussing good self-defense actions users tended to pro-actively list things they had not done (e.g. not clicked, not opened, not responded, no data entered) sometimes also adding that they had deleted or "junked" the email. These statements appeared to be an attempt to tell the help desk that the user was fine and did not require additional assistance from them. The following is a representative example of a user listing the actions they have not taken:

> I've attached a phishing email I received this morning. I didn't click any links and will now delete it from my inbox.

Or a similar example: "No links followed, no data entered, email being forwarded for information."

People sometimes described the steps taken which involved clicking a link before realizing that the email might be illegitimate. One example is as follows:

> I was suspicious, but it looks like a genuine [University email] address. However, the link did not take me to the university's OneDrive, so I closed it down immediately. Have I been stupid in clicking on this?

While these types of reports contain problematic actions the user took they also highlight positive actions such as noticing an issue and not proceeding. For example: "I received an email from WesternUnion when I clicked the link to pay; I have never received any communications from WesternUnion before and I am aware that they are often used for fraudulent purposes."

Finally, some users engaged with the email and were informing the help desk about the actions they took and sometimes also expressing uncertainty about if their actions resulted in anything problematic. Many such reports only contain the phishing message and a statement about what happened; for

example: "I unfortunately thought this was genuine and clicked on email and possibly entered details." Sometimes users have a hopeful tone that maybe the slip resulted in no real problem: "I did click on the open button but nothing came up." Or "I received this email today which I thought was from one of my colleagues [Name], when I clicked on the link [...] my computer started playing up, I closed it and it seems fine now." While the advice "change your password" may seem simple, for some users it was clearly a hassle so they wanted to be sure it was really necessary, as shown in the following example:

> Does this look like spam to you? I clicked the link and entered my password but it then asked me to reset my password - which I didn't do. If it is spam - should I change my usual password with whatever accounts I use it for (personal and work) – or do you think I'll be okay?

F. Questions (33, 12.2%)

All reports have the explicit or implied question of "is this email phishing?" so the *questions* codes were only applied when a user asked a further question. The most common question was asking the help desk for the *next steps* they should take. These queries ranged from a simple inquiry about if they can or should do anything (e.g. "Is there anything I should do?"), to a detailed description of potential actions and a request if they should be taken (e.g. "My query is whether I should contact [the people being impersonated] to let them know or just delete it and not waste any more time on it.").

The other questions and requests posed in phishing reports had a wide range of variation. Some users were looking for "how to" type answers often around protective actions. These included things like how to change their password, update their computer, or run a virus check. While the standard solution text provides direction on how to do these things, users had questions around non-standard situations. For example, one user changed the password they use to login to University websites but was confused if their email password was different or not. Another ran a virus check, then noted that the University guidance suggested a different way to run it that did not work, so inquired if the help desk could run it for them to be sure. People also wanted instruction on how to block such email in the future or asked the help desk if they could do so on their behalf.

Phishing email also caused some side effects that users were unsure how to undo. There was also a general concern around how to do common actions "safely" suggesting that users knew how to do these actions but were unsure how something caused by a malicious email might react to a common action. For example:

> I have received an Outlook calendar invitation from an unknown source, and I think it might be a phishing attempt. I want to remove it from my calendar. Please can you advise how I can do so safely. I deleted the invitation, but it is still showing as a recurring appointment in my calendar.

People were also unsure about how dangerous phishing email might be to the help desk staff and therefore inquired if it was safe to forward the email to them at all. For example, this user sent a screenshot of the email and then asked: "Please let me know if you would like me to forward the spam email I received if you would like to look into it further."

V. Discussion

Organizations put together phishing reporting structures with the aim of identifying phishing that makes it through the filters so it can be blocked and users protected, but an impressive 57.0% of reports contained statements or questions beyond simply reporting. Wash et al. [68] suggest that training should teach users that they are experts on the normal content of their own inboxes and that they should leverage this expertise. Our results confirm that some users are already thinking in this way. Particularly when presenting *evidence*, users discussed if they were expecting an email from that person or on that topic. They also called out cases where they were familiar with a service but found this type of contact abnormal. Interestingly, users are not just privately thinking through this logic, they are also attempting to convey it to help desk staff with the clear intention of having their evidence read and responded to.

When faced with a phishing message some users contacted the help desk because they honestly did not know how to handle the situation and were seeking guidance. Standard phishing guidance recommends things like deleting the email, changing the password, and running a virus scanner [42], these are all good advice but do not necessarily match the way users were expressing concerns. Phishing email often threatens the user with things like fees or having their accounts deleted; users contacted the help desk to determine if these threats were real and were therefore unwilling to follow advice to delete the email till they were assured. Users also have a poor understanding of how email technically functions leading to concerns around if different attacks were possible. For example, phishers would construct a fake email bounce notification which would cause users to think that their email address was being used to send messages to other people. They would then contact the help desk to figure out how to protect their account.

a) Standard solutions: The University help desk is charged with answering all sorts of questions and are not necessarily experts on phishing, so they worked with the University's security team to create a set of pre-written standard solutions around phishing queries. These solutions are intended to be technically accurate and target issues that the help desk sees regularly as well as issues the security team thinks are most important. All the solutions start with a clear statement about if the email is or is not malicious followed by advice and instructions. Common elements include telling users to avoid clicking and delete the email if they are unsure, how to reset a password, how to run an anti-virus, and how to regain access to a locked account.

Notably, the most common *action taken* described by users is a statement that they did not click anything, and the third is a statement that they have deleted the email. It is hard to say if the proactive statements are due to the standard solution language, but it is interesting that users felt the need to tell the help desk that they had taken good actions. Similarly, the second most commonly stated action was having clicked a link, which similarly suggests that users understand how problematic clicking can be.

Overall the standard solution approach was probably helpful to many but not all the reporters. Most users wanted to know if the email was or was not phishing, which the solution clearly states. Many were also looking for clear next steps or assurance that they had done the right thing. For common cases, like resetting a password or scanning for viruses, the standard solution also answers these questions. The solution likely matched less well in uncommon cases specific to the email being reported, such as the user receiving a threat and seeking assurance that the threat will not happen. Similarly, for cases where no protection action is possible, such as stopping an attacker from spoofing the user's email address.

b) Similarities with spotting phish: Prior work has looked at how people become aware of and identify phishing emails. In this work we observed that users put great focus on the *From address* of potential phishing emails, similar to what has been found by Downs et al. [19] and Beneson et al. [9] who also observed users paying attention to the from address or sender's identity. In our data, users mostly focused on if they knew the sender, recognized them, expected communication from them, or if the sender was using an email associated with the University. While from addresses are possible for attackers to spoof, most people cited instances of the from address being something other than what they expected. It is unclear if spoofed email addresses are uncommon in email that makes it through the automated filters, or if users were just better at recognizing and reporting non-spoofed email.

Similar to work by Wash et al. [68], we noticed that some users did engage in a type of investigation and reasoning to determine if the email was phishing before contacting the help desk. Most of the codes in *evidence* and *impact potential* are examples of users explaining their thinking to help desk staff. These explanations match well with Wash's stages of decisions [67], [68] where users collect evidence as well as the stage where they try and make decisions. Some users were clearly contacting the help desk for confirmation, laying out their reasoning and asking if the reasoning was correct or if the help desk had more information than they did.

c) Self-efficacy: The field of security has a long history of blaming users for poor security practices [1], [58], [71], users also have a tendency to blame themselves for making errors when it is actually poor usability that is at fault [45]. While coding the researchers engaged in memoing to capture observations. One such observation was around the language being used to self-describe actions they now knew to be wrong. Phrases like "I stupidly clicked" were common. Users also used self-derogatory language to justify why they were contacting the help desk for confirmation about if the email was phishing or not rather than sorting it out on their own. One person even self-tagged their message "tinfoilhat" presumably to convey that they knew they might be being overly cautious, even though the phishing they were contacting about was indeed real. Prior work has observed that users sometimes treat those who are overly cautious security wise as "paranoid" [24] and our results suggest that users may view reporting phishing in a similar light. Such a situation is problematic, because it means that users may be avoiding reporting phish in an effort to not be seen as "paranoid". This lack of self-efficacy is potentially similar to the case of privacy management, where we can see potential benefits and costs dependent on the individuals confidence around their choices [14].

Through encouragement of reporters, in the form of feedback on reports, we could improve individuals' self-efficacy around spotting and reporting phish. Positive feedback for phishing cases could potentially educate users and make them more aware of their untapped potential. Having engaged and willing end-users will remain a key component in an organization's defence against phishing attacks, and current means do not do enough to incorporate and show appreciation for the efforts of reporters [4].

A. Future Directions

Based on our observations, we highlight four main areas to address to strengthen the phishing reporting processes in the organizations. Our findings also show that more work needs to be done to understand users better and make the phishing reporting processes more inclusive.

a) Phishing education: Our findings highlight that many of our users reporting phish had their suspicions raised due to contradictions from the sender's email address. For example, the attackers are often using common mail services, such as Outlook or Gmail, and yet claiming to be from say 'HMRC' (the tax authority of the UK). This shows that topics like the from address make sense to people and possibly should be emphasized more in training, and instead of focusing training on features that users do not understand, such as email headers [70] or URL parsing [2]. Furthermore, future research should identify what other features users actually understand and are able to identify as ways of grounding their investigations of suspicious emails and potential phish while utilising their unique contextual understanding of their email inboxes [68]. Our work highlights some potential contextual factors which users respond to when identifying phishing, and we should therefore adapt current messaging and training on phishing scams to focus on what users appear to already be successful with.

b) Encouraging phishing reporting: A common theme that we found was that end-users would only report phish that were particularly hard to spot [38]. Although helpful, organizations and their IT staff would prefer to have any and all examples of phish reported to them as these examples can be used to improve the technical barriers of their systems, preventing future examples from reaching the inboxes of their staff and colleagues. In a similar vein, we saw that users would be selective to avoid "wasting IT teams' time", which is thoughtful but also detrimental to the overall security of an organization as it means that emails users judge as obvious may not be removed or blocked, especially when such emails are the ones that could not be identified by sophisticated AI-based filters. Work must continue to identify how best to report phish such that IT staff are not over run with reports [4], and are provided with a representative sample of the ongoing attacks their organization faces.

c) Reporting systems and Human-AI Collaboration: In current practices, the reporting process can be cumbersome to go through for some users. This highlights a need for better designed services: (i) to help the users to report without discouraging future reporting, (ii) to gain the trust of the users by providing customized feedback beyond the generic

guidance [31]. Using non-judgmental language within the feedback provided would also endear continual reporting. For example, a visible button could be deployed on the email clients, or the phishing feedback itself could be done in a more contextualized manner. AI could also be useful to make sense of the content of the emails [56]. For example, large language models could be used to extract contextual features, and natural language processing techniques could be applied to automatically map features in reported phish to the types of information identified in our codebook. In other words, a hybrid approach of humans and AI could provide a digestible format of the reports in a structured manner. This would also ease the report handling process by the IT staff. As shown in Section III-C, half of the reported emails included non-descriptive subject lines. Hence, the IT staff also need better tools to understand the nature of the phishing emails being reported and to manage them efficiently by providing ranges of technical features which can be used to improve systems and their defences [4].

d) Reassuring users: Our work shows that users are often looking for reassurance rather than guidance. Sometimes the reassurance is easy to provide, such as assuring them that the email is phishing and they did the right thing reporting it. But other assurance is harder due to the time required to read and respond to reports or because the question involves organizations other than the University. For example, users want to know if they have taken the right steps. For common cases the standard solution response may be enough, but users also want to know if it is fine not to take action. Such as not changing their password because they "just clicked" and did not give the password away. Or if it is also necessary to report to the police, or if reporting to the help desk is enough. These cases are harder to answer quickly. Future research should consider how to address these types of questions at scale taking into account the time limitations of a help desk. There is some potential for AI techniques in generating more context-specific advice that the help desk staff could then quickly use and, then customize it even further according to the user requests [31]. There may also be a better set of guidance to put in standard solutions or FAQ pages that helps cover some of these issues.

VI. Conclusions

In conclusion, we qualitatively analyzed a random sample of 270 reported phishing tickets submitted over approximately nine months. The reports were submitted by a wide range of people associated with the University with most people reporting only one or two times within the time frame. Through the detailed qualitative analysis we find that 57.0% of the tickets contain additional information or questions. Users provided *evidence* to explain why they think the reported email is phishing, they also explain why they feel the email has *impact potential*, what *actions* they have or have not taken, and also ask a range of *questions* to the help desk. Users were often looking for next steps, guidance, or reassurance that their judgement or actions were appropriate. They were also trying to explain why they thought that the email was phishing, or why they thought this email in particular was dangerous and deserved attention.

The results suggest that at least among reporters, users have a good sense of common phishing cues and what to avoid doing with an identified phish, such as not clicking anything. In terms of organizational protection, it also suggests that users are not forwarding all phish that they see and instead only reporting exceptional phishing examples that they are uncertain about. Future research should look at users' mental models around reporting phishing to understand what users expect from reporting phish and how to support them so that organizations can benefit from more reporting.

Acknowledgment

We would like to thank the reviewers and members of the TULiPS lab for their feedback on this work. Additionally, we thank our collaborators in the IT department of the University. This research is supported by REPHRAIN: The National Research Centre on Privacy, Harm Reduction and Adversarial Influence Online, under UKRI grant: EP/V011189/1,

References

[1] A. Adams and M. A. Sasse, "Users are not the enemy," in *Communications of the ACM*, 1999.

[2] S. Albakry, K. Vaniea, and M. K. Wolters, "What is this URL's destination? Empirical evaluation of users' url reading," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–12. [Online]. Available: https://doi.org/10.1145/3313831.3376168

[3] M. Alsharnouby, F. Alaca, and S. Chiasson, "Why phishing still works: User strategies for combating phishing attacks," *International Journal of Human-Computer Studies*, vol. 82, pp. 69–82, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581915000993

[4] K. Althobaiti, A. D. G. Jenkins, and K. Vaniea, "A case study of phishing incident response in an educational organization," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW2, oct 2021. [Online]. Available: https://doi.org/10.1145/3476079

[5] K. Althobaiti, N. Meng, and K. Vaniea, "I don't need an expert! Making URL phishing features human comprehensible," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, May 2021, pp. 1–17. [Online]. Available: https://groups.inf.ed.ac.uk/tulips/papers/althobaiti2021chi.pdf

[6] K. Althobaiti, G. Rummani, and K. Vaniea, "A review of human-and computer-facing url phishing features," in *IEEE European Symposium on Security and Privacy Workshops (EuroSPW)*. IEEE, 2019, pp. 182–191. [Online]. Available: https://groups.inf.ed.ac.uk/tulips/papers/althobaiti2019.pdf

[7] K. Althobaiti, K. Vaniea, and S. Zheng, "Faheem: Explaining URLs to people using a slack bot," in *2018 Symposium on Digital Behaviour Intervention for Cyber Security (AISB)*. Liverpool, UK: University of Liverpool, Apr. 2018, pp. 1–8. [Online]. Available: http://aisb2018.csc.liv.ac.uk/PROCEEDINGS\%20AISB2018/Digital\%20Behaviour\%20Interventions\%20for\%20CyberSecurity\%20-\%20AISB2018.pdf

[8] N. A. G. Arachchilage, S. Love, and K. Beznosov, "Phishing threat avoidance behaviour: An empirical investigation," *Computers in Human Behavior*, vol. 60, pp. 185–197, 2016. [Online]. Available: https://doi.org/10.1016/j.chb.2016.02.065

[9] Z. Benenson, F. Gassmann, and R. Landwirth, "Unpacking spear phishing susceptibility," in *Financial Cryptography and Data Security*, M. Brenner, K. Rohloff, J. Bonneau, A. Miller, P. Y. Ryan, V. Teague, A. Bracciali, M. Sala, F. Pintore, and M. Jakobsson, Eds. Cham: Springer International Publishing, 2017, pp. 610–627.

[10] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All your contacts are belong to us: automated identity theft attacks on social networks," in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 551–560.

[11] M. Blythe, H. Petrie, and J. A. Clark, "F for fake: Four studies on how we fall for phish," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 3469–3478. [Online]. Available: https://doi.org/10.1145/1978942.1979459

[12] G. Canova, M. Volkamer, C. Bergmann, and B. Reinheimer, "Nophish app evaluation: lab and retention study," in *NDSS workshop on usable security*, 2015.

[13] Y. Cao, W. Han, and Y. Le, "Anti-phishing based on automated individual white-list," in *Proceedings of the 4th ACM workshop on Digital identity management*, 2008, pp. 51–60.

[14] H.-T. Chen and W. Chen, "Couldn't or wouldn't? the influence of privacy concerns and self-efficacy in privacy management on privacy protection," *Cyberpsychology, behavior and social networking*, vol. 18 1, pp. 13–9, 2015.

[15] A. Cidon, L. Gavish, I. Bleier, N. Korshun, M. Schweighauser, and A. Tsitkin, "High precision detection of business email compromise," in *28th USENIX Security Symposium (USENIX Security 19)*. Santa Clara, CA: USENIX Association, Aug. 2019, pp. 1291–1307. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/cidon

[16] D. L. Cook, V. K. Gurbani, and M. Daniluk, "Phishwish: a simple and stateless phishing filter," *Security and Communication Networks*, vol. 2, no. 1, pp. 29–43, 2009. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sec.45

[17] S. Das, C. Nippert-Eng, and L. J. Camp, "Evaluating user susceptibility to phishing attacks," *Information & Computer Security*, 2022.

[18] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006, pp. 581–590.

[19] J. S. Downs, M. B. Holbrook, and L. F. Cranor, "Decision strategies and susceptibility to phishing," in *Proceedings of the second symposium on Usable privacy and security*, 2006, pp. 79–90.

[20] I. Fette, N. Sadeh, and A. Tomasic, "Learning to detect phishing emails," in *Proceedings of the 16th International Conference on World Wide Web*, 2007, pp. 649–656.

[21] U. Finance, "Fraud - The Facts 2021, The Definitive Overview of Payment Industry Fraud," 2021, also available as https://www.ukfinance.org.uk/system/files/Fraud\%20The\%20Facts\%202021-\%20FINAL.pdf. Accessed Feb. 2022.

[22] A. Franz, V. Zimmermann, G. Albrecht, K. Hartwig, C. Reuter, A. Benlian, and J. Vogt, "SoK: Still plenty of phish in the sea — a taxonomy of User-Oriented phishing interventions and avenues for future research," in *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*. USENIX Association, Aug. 2021, pp. 339–358. [Online]. Available: https://www.usenix.org/conference/soups2021/presentation/franz

[23] E. D. Frauenstein and R. von Solms, "An enterprise anti-phishing framework," in *Information Assurance and Security Education and Training*, R. C. Dodge and L. Futcher, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 196–203.

[24] S. Gaw, E. W. Felten, and P. Fernandez-Kelly, "Secrecy, flagging, and paranoia: adoption criteria in encrypted email," in *CHI '06: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 2006, p. 591–600.

[25] A. F. Hayes and K. Krippendorff, "Answering the call for a standard reliability measure for coding data," *Communication methods and measures*, vol. 1, no. 1, pp. 77–89, 2007.

[26] "Report phishing sites," https://www.us-cert.gov/report-phishing, 2020, accessed Feb. 2022.

[27] J. Hong, "The state of phishing attacks," *Commun. ACM*, vol. 55, no. 1, p. 74–81, jan 2012. [Online]. Available: https://doi.org/10.1145/2063176.2063197

[28] P. Jaccard, "Nouvelles recherches sur la distribution florale," *Bulletin de la Societe vaudoise des Sciences Naturelles*, vol. 44, pp. 223–270, 1908.

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

[30] M. Jakobsson, A. Tsow, A. Shah, E. Blevis, and Y. kyung Lim, "What instills trust? a qualitative study of phishing," in *Proceedings of the 11th International Conference on Financial Cryptography and 1st International Conference on Usable Security*, 2007, p. 356–361.

[31] A. Jenkins, N. Kokciyan, and K. E. Vaniea, "Phished: Automated contextual feedback for reported phishing," in *18th Symposium on Usable Privacy and Security*. Usenix, 2022.

[32] K. D. Joshi and P. Nalwade, "Modified k-means for better initial cluster centres," *International Journal of Computer Science and Mobile Computing*, vol. 2, no. 7, pp. 219–223, 2013.

[33] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: A literature survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 2091–2121, 2013.

[34] K. Krippendorff, *Content analysis: An introduction to its methodology*. Sage publications, 2018.

[35] P. Kumaraguru, J. Cranshaw, A. Acquisti, L. Cranor, J. Hong, M. A. Blair, and T. Pham, "School of phish: A real-world evaluation of anti-phishing training," in *Proceedings of the 5th Symposium on Usable Privacy and Security*, ser. SOUPS '09. New York, NY, USA: Association for Computing Machinery, 2009. [Online]. Available: https://doi.org/10.1145/1572532.1572536

[36] P. Kumaraguru, Y. Rhee, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge, "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*, 2007, pp. 905–914.

[37] P. Kumaraguru, S. Sheng, A. Acquisti, L. F. Cranor, and J. Hong, "Teaching johnny not to fall for phish," *ACM Trans. Internet Technol.*, vol. 10, no. 2, jun 2010. [Online]. Available: https://doi.org/10.1145/1754393.1754396

[38] Y. Kwak, S. Lee, A. Damiano, and A. Vishwanath, "Why do users not report spear phishing emails?" *Telematics Informatics*, vol. 48, p. 101343, 2020. [Online]. Available: https://doi.org/10.1016/j.tele.2020.101343

[39] T. Lin, D. E. Capecci, D. M. Ellis, H. A. Rocha, S. Dommaraju, D. S. Oliveira, and N. C. Ebner, "Susceptibility to spear-phishing emails: Effects of internet user demographics and email content," *ACM Trans. Comput.-Hum. Interact.*, vol. 26, no. 5, jul 2019. [Online]. Available: https://doi.org/10.1145/3336141

[40] J. MacQueen, "Classification and analysis of multivariate observations," in *5th Berkeley Symp. Math. Statist. Probability*, 1967, pp. 281–297.

[41] K. A. Moul, "Avoid phishing traps," in *ACM SIGUCCS Annual Conference, SIGUCCS*. New Orleans, LA, USA: ACM, 2019, pp. 199–208. [Online]. Available: https://doi.org/10.1145/3347709.3347774

[42] "Phishing attacks: dealing with suspicious emails and messages," http://bit.ly/3tTwQpC, Dec. 2018, accessed Feb. 2022.

[43] L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguyen, "A novel approach for phishing detection using url-based heuristic," in *2014 International Conference on Computing, Management and Telecommunications (ComManTel)*, 2014, pp. 298–303.

[44] J. Nicholson, L. Coventry, and P. Briggs, "Can we fight social engineering attacks by social means? assessing social salience as a means to improve phish detection," in *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. Santa Clara, CA: USENIX Association, Jul. 2017, pp. 285–298. [Online]. Available: https://www.usenix.org/conference/soups2017/technical-sessions/presentation/nicholson

[45] D. Norman, *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[46] D. Oliveira, H. Rocha, H. Yang, D. Ellis, S. Dommaraju, M. Muradoglu, D. Weir, A. Soliman, T. Lin, and N. Ebner, "Dissecting spear phishing emails for older vs young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 6412–6424. [Online]. Available: https://doi.org/10.1145/3025453.3025831

[47] J. Petelka, Y. Zou, and F. Schaub, "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*, 2019, pp. 1–15.

[48] A. Rajaraman and J. D. Ullman, *Data Mining*. Cambridge University Press, 2011, p. 1–17.

[49] J. Ramos *et al.*, "Using tf-idf to determine word relevance in document queries," in *Proceedings of the first instructional conference on machine learning*, vol. 242, no. 1. New Jersey, USA, 2003, pp. 29–48.

[50] D. Rathee and S. Mann, "Detection of e-mail phishing attacks - using machine learning and deep learning," *International Journal of Computer Applications*, vol. 183, pp. 1–7, 01 2022.

[51] E. M. Redmiles, S. Kross, and M. L. Mazurek, "How I learned to be secure: a census-representative survey of security advice sources and behavior," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, October 24-28, 2016*. Vienna, Austria: ACM, 2016, pp. 666–677. [Online]. Available: https://doi.org/10.1145/2976749.2978307

[52] B. Reinheimer, L. Aldag, P. Mayer, M. Mossano, R. Duezguen, B. Lofthouse, T. Von Landesberger, and M. Volkamer, "An investigation of phishing awareness and education over time: When and how to best remind users," in *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security*, ser. SOUPS'20. USA: USENIX Association, 2020.

[53] J. Reynolds, D. Kumar, Z. Ma, R. Subramanian, M. Wu, M. Shelton, J. Mason, E. Stark, and M. Bailey, "Measuring identity confusion with uniform resource locators," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–12. [Online]. Available: https://doi.org/10.1145/3313831.3376298

[54] T. Ronda, S. Saroiu, and A. Wolman, "Itrustpage: A user-assisted anti-phishing tool," in *Proceedings of the 3rd ACM SIGOPS/EuroSys European Conference on Computer Systems 2008*, ser. Eurosys '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 261–272. [Online]. Available: https://doi.org/10.1145/1352592.1352620

[55] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from urls," *Expert Systems with Applications*, vol. 117, pp. 345–357, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417418306067

[56] T. Saka, K. Vaniea, and N. Kökciyan, "Context-based clustering to mitigate phishing attacks," in *Proceedings of the 15th ACM Workshop on Artificial Intelligence and Security*, ser. AISec'22. New York, NY, USA: Association for Computing Machinery, 2022, p. 115–126. [Online]. Available: https://doi.org/10.1145/3560830.3563728

[57] J. Saldaña, *The coding manual for qualitative researchers*. Sage, 2015.

[58] M. A. Sasse, M. Smith, C. Herley, H. Lipford, and K. Vaniea, "Debunking security-usability tradeoff myths," *IEEE Security and Privacy*, vol. 14, no. 5, pp. 33–39, 2016. [Online]. Available: http://ieeexplore.ieee.org/abstract/document/7676175/

[59] T. Seals, "Cost of user security training tops $290k per year." 2017, also available as https://www.infosecurity-magazine.com/news/cost-of-user-security-training. Accessed Nov. 2020.

[60] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs, "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*, 2010, pp. 373–382.

[61] S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. I. Hong, and E. Nunge, "Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish," in *Proceedings of the 3rd Symposium on USAble Privacy and Security, SOUPS*, vol. 229. Pittsburgh, Pennsylvania, USA: ACM, Jul. 2007, pp. 88–99. [Online]. Available: https://doi.org/10.1145/1280680.1280692

[62] S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang, "An empirical analysis of phishing blacklists," 2009.

[63] G. Stringhini and O. Thonnard, "That ain't you: Blocking spearphishing through behavioral modelling," in *Detection of Intrusions and Malware, and Vulnerability Assessment*, M. Almgren, V. Gulisano, and F. Maggi, Eds. Cham: Springer International Publishing, 2015, pp. 78–97.

[64] Verizon, "2019 data enterprise phishing resiliency and defense report breach investigations report," Verizon Trademark Services LLC, Tech. Rep., 2019, also available as https://enterprise.verizon.com/resources/reports/2019-data-breach-investigations-report.pdf. Accessed Jun. 2020.

[65] ——, "2021 data breach investigations report," Verizon Trademark Services LLC, Tech. Rep., 2021, also available as https://www.verizon.com/business/resources/reports/dbir/?CMP=OOH_SMB_OTH_22222_MC_20200501_NA_NM20200079_00001. Accessed February. 2022.

[66] M. Volkamer, K. Renaud, B. Reinheimer, and A. Kunz, "User experiences of Torpedo: Tooltip-powered phishing email detection," *Computers & Security*, vol. 71, pp. 100–113, 2017.

[67] R. Wash, "How experts detect phishing scam emails," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–28, 2020.

[68] R. Wash, N. Nthala, and E. Rader, "Knowledge and capabilities that {Non-Expert} users bring to phishing detection," in *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, 2021, pp. 377–396.

[69] W. Yang, A. Xiong, J. Chen, R. W. Proctor, and N. Li, "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*. Hanover, MD, USA: ACM, Apr. 2017, pp. 52–61. [Online]. Available: https://doi.org/10.1145/3055305.3055310

[70] S. Zheng and I. Becker, "Presenting suspicious details in User-Facing e-mail headers does not improve phishing detection," in *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. Boston, MA: USENIX Association, Aug. 2022, pp. 253–271. [Online]. Available: https://www.usenix.org/conference/soups2022/presentation/zheng

[71] M. E. Zurko and R. T. Simon, "User-centered security," in *Proceedings of the 1996 workshop on New security paradigms*, 1996, pp. 27–33.